

STIC-Biotech/ChemLib

58170

From: Walicka, Malgorzata  
Sent: Monday, January 14, 2002 10:53 AM  
To: STIC-Biotech/ChemLib

I would appreciate the search for:

SEQ ID NO: 1 and 2. Application No. 09/833,782, the earliest priority is 4/12/2000. Please use the pending and commercial data bases.  
Please RUSH!

Thank you,  
Malgorzata Walicka  
Art Unit 1652  
tel.305-7270  
Mali box 10C01  
Room 10D06

Point of Contact:  
Mona Smith  
Info. Specialist  
GM1 12314 Tel: 800-3273

CRFE

STIC-Biotech/ChemLib

From: Chan, Christina  
Sent: Monday, January 14, 2002 11:53 AM  
To: STIC-Biotech/ChemLib; Walicka, Malgorzata  
Subject: FW:

**Please rush. Thanks Chris**

-----Original Message-----

From: Walicka, Malgorzata  
Sent: Monday, January 14, 2002 10:57 AM  
To: Chan, Christina  
Cc: STIC-Biotech/ChemLib  
Subject:

Christina, could you authorize this rush search?

Thank you,

Malgorzata

Searcher: M. Smith  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 11/4/02  
Date Completed: 11/5  
Searcher Prep/Review: 10  
Clerical: \_\_\_\_\_  
Online time: 10

TYPE OF SEARCH:

NA Sequences: 1  
AA Sequences: 1  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST(where applic.)

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

---



**THIS PAGE BLANK (USPTO)**



/translation="MTLGRFMSPLQAMSSYVDGRNVLRLWDLSPQIKRRTEELIAQ  
TKQVYDGLMIEEVTYENCLOALADVEVKYIIVERTMLDFPOHVSSDKREVAATEA  
DKRLSRFLEMSEMDIFURIVLKEDETLGKIKPARYLEKSKPMKGRNGHLPQEO  
VNEIKRMDKRRSELCIDFNKLNEDDETFLVSKAELGALPDDFIDLSLEKTDNNKYKI  
TLKYPHYFVPMKCCIPETRRKWEAFNTRCKEENTIILOELLPLRAKVLKLGSTH  
ADFLVEMTAKSTHVAFLDLDLSOKLPLGEAREFIILNKKKEEKFEVGDGIN  
ANDLHYMTOTBELKYSVDQELKVEFLEVTUEGLLNYQELLGLSFEQVVDARVWN  
KSTLYTVKDKATGEVILGOPYLDLYPRECKYNNHACFGLOPCLLPDSRQRMWSVAALV  
VNFSPRAGRPSLLRDEVRTYFHEFGVHMOICATDFARFSGTNVDFEYFVSOM  
LENVWMDTSLRRLSKHYKDGSPITDDLLKLVASRLVNTGLLTROIIVLSVQDLSLH  
TNTSDAADESKAYCTEILGVAATPTGNPATFGLIAGGYDGYTYGLWSEVFSMDMF  
YSCFKKEGIMNPEVGMKYNLILKPGSLDGMQLNQLFKREPKNQKFLMSRLHAP"  
BASE COUNT 826 a 545 c 641 g 720 t  
ORIGIN

Query Match 87.9%; Score 1858.2; DB 4; Length 2732;  
Best Local Similarity 92.3%; Pred. No. 0;  
Matches 1953; Conservative 2; Mismatches 160; Indels 0; Gaps 0;  
QY 1 atgacgcccgggtgccttttggtgctgcgaagcctccgcagagttggtgtccaggatt 60  
DB 175 ATGATGCTCGGTGCTTTTCGGTGGGAGAGAAATGTCCTCTTCAGGCAATGTCCTCTAT 234  
QY 61 ttactcagaatcagcttagaagagagtgatgtctcctcttcaggcaatgtcttccatt 120  
DB 235 TTATTCAGAGATGACTTTAGGAGAGAGAAATGTCCTCTTCAGGCAATGTCCTCTAT 294  
QY 121 actgtgctggcagaaatgftttaagatggatctttaccagagctcgcagagttggtgtccaggatt 180  
DB 295 ACTGTGGATGGCAGAAATGTTTAAAGTGGGATCTTTACACAGAGCAATTAAGAAAGA 354  
QY 181 actgaggagctcattgtgcagaccacacaggtgtacgtgtctgttggaatgctcgtgatt 240  
DB 355 ACTGAGGAGCTGATTCGCAGACCAACAGGCTATGATGATATGGAATGCTTGACATT 414  
QY 241 gaggaagtaactacagagacgtctgcaggcactggcagatgtagaagttaagtata 300  
DB 415 GAAGAGGTAACTTATGAGAACTGTTTGCAGGACTGGCAGATGTGGAAGTGAAGTATATA 474  
QY 301 gtggaagagaccatgtagacttccacagcatgtatcctctgcagaaagaaagtagagca 360  
DB 475 GTGGAACAGACCATGCTAGACTTTCCCAAGCATGCTCCTCTGACAAAGATACAGACA 534  
QY 361 gcaagtacagaagcagacaaaagactttctgtttttgattgagatgagcatgagagga 420  
DB 535 GCAAGTACAGAAGCAGACAAAAGACTTTCTCGTTTGTATTTGATGATGAGATGAGATGAGAA 594  
QY 421 gatatttgagagagattgttcatttacaggaaacactgtatctctggggaagataaacct 480  
DB 595 GATATCTTTCTGAGAAATGTTCTGTTAAAGAGAGACTGTGATTTGGGGAAGATCAAGCCT 654  
QY 481 gaggcagacgatacttgaaagcaatcaattaaatggggaagaaagaaatgggtccattctt 540  
DB 655 GAGCCAGAGAGATCTTGAAGAAATCAGTTAAATGGGAAAGAAATGGGCTCCATCTT 714  
QY 541 cctgacaagtagacagaatgaatcaaatcaatgaagaaagaaatgagtgatgtatt 600  
DB 715 CGGGAACAAGTACAGAAATCAAGCAATGAGAAAGAAATGAGTGAGGTATGTATT 774  
QY 601 gattttaaaaaaacctcaatgagatgataccttctgtatttttccaggctgaact 660  
DB 775 GACTTTCAACAAAACCTCAATGAGGACGATACCTTCTTGTATTTTCCAAAGGCTGAACCT 834  
QY 661 ggtgctctcctgatttcaatgacagtttagaagacagatgagcaagtagataaa 720  
DB 835 GGTGCTCTCTCTGATGATTCTATTCACAGTTTTAGAAAAAGACAGATGCAACAAGTATATA 894  
QY 721 attaccttaaatatccacactatttccctgtcatgaagaaatgtgtatcctcctgaacc 780  
DB 895 ATTACCTTAAAAATATCCACACTATTTTCTGTCATGAGAAATGTTGTATTCCTTGAAACC 954  
QY 781 agaagagagtggaatggccttttaatacaagggtgcgaagggaaacaccacataattttg 840

DB 955 AGAAGAAAATGAAATGGCTTTTAATACAGGTGCAAGAGGAAACACCAATAATCTCTG 1014  
QY 841 cagcactactccactcgcgaacaaaggtggcaaaactactcgtgttatagcacacatgct 900  
DB 1015 CAGAGCTACTCCACTCGGGCAAAGTAGCCAAACTACTGGGCTACAGCACACATGCT 1074  
QY 901 gacttcgcttgtaaatgaactgcacagagcacaagcgcgcgttaacagccttcttagat 960  
DB 1075 GACTTTGTCTTCCTTGAGATGAACACTGCGAAGAGTACACACCATGTGACAGCCTTTCAGAT 1134  
QY 961 gatttaagcagaagttaaaaccccttggtggaagcagaacagagagttatttttgaatttg 1020  
DB 1135 GATTTAAGCCAGAAATTAATAACCTTGGGTGAGGCAAGACGAGAGTTCAATTTTGAATTTG 1194  
QY 1021 aagaaaaaggaatgcacagacaggggttttgaatatgatgggaaaaatcaatgcctgggat 1080  
DB 1195 AAGAAAAAGGAATGTGAAGAGAAAGGTTTGAATATGATGGCAAAATCAATGCTGGGAT 1254  
QY 1081 ctatatctactactgactcagacagaggaactcaagtatttccatagaccaagagttctctc 1140  
DB 1255 CTACATTACTATGACTCAGACAGAGAACTCAAGTACTCTGTGGACCAAGAGATCTCTC 1314  
QY 1141 aaggaatacttcccaattgaggtgctcactgaagcgttgcacacacctaccagagagttg 1200  
DB 1315 AAGGAATACTTCCCAATTTGAAGTGGTCACTGAAGGCTTCTCTGAACATCTATCAGGAGTTG 1374  
QY 1201 ttgggaacttctatgaaacaaatgacagatgctcatgttttggaaacaaagagtttcaactt 1260  
DB 1375 TTGGGACTCTCTCTTTCAGCAAGTACAGATGCTCATGTTTGAATAAATAAGTGTTCACCTG 1434  
QY 1261 tatactgtgaagataaagctacagagaagatttgggacagttcttatttggacctctat 1320  
DB 1435 TACACGGTGAAGGATAAAGCCCAAGAGAAAGTATTAGGACAGTTTCTACTTGGACCTCTAT 1494  
QY 1321 ccaaggggaagaaaaatacaatcatcgccgtcctcgtcctccacccgctggcgtcctctg 1380  
DB 1495 CCAAGGGAAGGAAATACAACTACGACCTGCTTGGTCTCCAGCCTGGCTGGCTCTCTG 1554  
QY 1381 cctgatgaaagccgagatgagcgagtggtgcctcctcgttggtaacttctcacagccagt 1440  
DB 1555 CCGTATGGAGAGCCGATGATGCTGTGGCTGCCCTCGTGGTGAACCTTCTCACAGCACGA 1614  
QY 1441 gcagctcgtccctctcctcagacacacagagtgagagacttaacttcaatgagtttgt 1500  
DB 1615 GCAGTGTGCTCTCTCTCTGAGACACGATGAGGTGAGGACTTACTTCCATGAATTTGCT 1674  
QY 1501 cactgtatgcatcagatttgcacagactgatttgcacagatttgcagcagatttagcggaacaaatgtg 1560  
DB 1675 CACGTCATGCATCAGATTGTGGCAGACTGACTTGTGACGATTTAGTGGAAACAATGTG 1734  
QY 1561 gaaactgactttgtagagtgccatcgcaaatgcttgaaaaattgggtgtggagcagctgatt 1620  
DB 1735 GAAACTGACTTTGTAGAGTGCATCACAAAATGCTTTGAAAAGTGGGTGGGACACTGAT 1794  
QY 1621 tccctccgaagatgtcaaaaactataaagatggaagccctattgcagagatctgctt 1680  
DB 1795 TCCCTCCGAAGGCTGTCAAAAACATATTAAAGATGGAAGCCCTATTACAGATGATCTGCTT 1854  
QY 1681 gaaaaactgttgcctcctaggtgtcaacacagctctctgacccctgcgcagagattgtt 1740  
DB 1855 GAAAAACTTGTGCTTCTAGACTGGTCAACACAGGCTCTCCGACCTTACGCCAGATGCTT 1914  
QY 1741 ttgagcaagttgacagttctcttcatacccaacacatcgtggtgagtgctgcaagtgaat 1800  
DB 1915 TTGAGCAAGTTGATCAGTCTCTCCACACCAACTTCCGCTGGATGCTGCAAGTGAATAT 1974  
QY 1801 gccaaatctcctcagaaataattaggaggttcagctactccagcacaataatgccagct 1860  
DB 1975 GCCAAATCTCAGCAAGAAATTTTAGGTGTTGACAGTCTACTCCAGGCACAAAATATGCCAGCT 2034  
QY 1861 accttggacatttggcaggggagatagatggcgaatattatgatcttcttggagtaa 1920

**THIS PAGE BLANK (USPTO)**

```

Db 2035 ACTTTGGACACTTGGCAGGGGATATGATGGCCAATATTATGATATCTTTGGAGTAA 2094
QY 1921 gtatttccatgatattttacagctgttttaaaagaaggataatgaatccagag 1980
Db 2095 GTCTTTCCATGACATGTTTATAGCTGTTTAAAGAAAGAGGATAATGAATCCAGAG 2154
QY 1981 gttggaatgaataacacgaacacctaactcgaacacctgggggagctctctgacgagcatggag 2040
Db 2155 GTTGAATGAATACAGAAACCTAATCTCTAAACCTGGGGGATCGCTGGAGCGCATGGAC 2214
QY 2041 atgctccacaatttcttgaacagtcgagcaaaccaaaaacgcttcttaataagtagagc 2100
Db 2215 ATGCTCCAAAATCTTGAACGTGAGCCAAACCAAAACCGTTCTCTAATGAGTAGAGGC 2274
QY 2101 ctgcattctctcgtga 2115
Db 2275 CTGATGCTCGGTAA 2289

RESIDUE 4
LOCUS PIGSABP 3819 bp mRNA MAM 14-APR-2000
DEFINITION Sus scrofa mRNA for soluble angiotensin-binding protein, complete cds.
ACCESSION D11336
VERSION D11336.1 GI:217708
KEYWORDS SINE; SAMP; short interspersed repeated sequence; soluble angiotensin-binding protein.
SOURCE Sus scrofa (sub-species:domesticus) heart, liver cDNA to mRNA, clone_l1b:lambdazap II clone:lambdapAB[1.4,s2,-L1].
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 3819)
AUTHORS Sugiura,N., Hagiwara,H. and Hirose,S.
TITLE Molecular cloning of porcine soluble angiotensin-binding protein
JOURNAL J. Biol. Chem. 267 (25), 18067-18072 (1992)
MEDLINE 9238174
REFERENCE 2 (bases 1 to 3819)
AUTHORS Hirose,S.
TITLE Direct Submission
JOURNAL Submitted (18-MAY-1992) to the DDBJ/EMBL/GenBank databases.
Shigehisa Hirose, Tokyo Institute of Technology, Department of Biological Sciences, Ookayama, Meguro-ku, Tokyo 152, Japan (Tel:03-3726-1111(ex.2226), Fax:03-3729-0335)
FEATURES
source
1..3819
/organism="Sus scrofa"
/sub_species="domesticus"
/db_xref="taxon:9823"
/clone="lambdapAB[1.4,s2,-L1]"
/clone_l1b="lambdazap II"
/tissue_type="heart, liver"
158..2272
/note="SAMP"
/codon_start=1
/product="soluble angiotensin-binding protein"
/protein_id="BA01949.1"
/db_xref="GI:217709"
CDS
/translation="MIVRCLSAARLHRVGGSGILLRMTLIGREAMSPLOAMSSYVDG
RNYLRWLSPEQIKRTEEDIAQTKQVDDIGMDIEEVTYENCLOALADVEVKYIVE
RTWLDPPQHVSKDEYRAASTADKRLSRFDIEMSMREDIFLRIVRKFTCDLGIKP
EARYLEKSVKMGKRNGLHPEQVNEIKAMKRMSELIDFNKNNLEDTHLVFSA
ELGALPDDFIDSLKEDDNKIKITLPIFYFVNNKCCIPETRRKMEMAFNTRCKEN
TIILQELLPUKAVKLLGYSTHADFLVEMNTAKSTHVTAFDDLSOKLPGEAEGR
EFTLNKKECEKEFGYDKINAWDLHYMTTEELKYSDQELIKYFPFIEVTFEG
LLNIYBLGLISPEQVDHAWNKNVLTIVKDKATGEVLQFYLILYREGYNHAA
CFGLPGCLLPDGSRRMVAALVNFVSOEAPRPSLLRHDEVTYFHEGHVHOICA
QTFAPFSGNVEFDFVPSOMLENWVDTSLRLSKHYKDGSPITDLDLKLIVAS
RLVNTGLLTURQIVLSKVDQSLHNTSLDASEYAKYCTEILGVAATPGTNMPATFGH
LAGGYDQYQYVLSVEYFSDMFYSCFKREGIMNPEVGMKRYRLILKPGSLDGMDL
QNFLKREPNGKAFILMSRLHAP"
2001..2137
variation

```

```

variation
2168..2169
/note="deleted from clone lambdapAB1"
/clone="substituted in clone lambdapAB1"
polyA_signal 2676..2681
polyA_signal 2989..2994
repeat_unit 3459..3780
/note="SINE(PRE-1)"
BASE COUNT 1086 a 788 c 855 g 1090 t
ORIGIN
Query Match 87.9%; Score 1858.2; DB 4; Length 3819;
Best Local Similarity 92.3%; Pred. No. 0;
Matches 1953; Conservative 2; Mismatches 160; Indels 0; Gaps 0;
QY 1 atgacgcgcgcgtcttttggctgtgcgaagccctccgcagagttggtgttccaggatt 60
Db 158 ATGATGTCCTCGGTGCTTTCGGCTCGGAGAGCCCTCCACAGGGTGGTGTCTCCGGATT 217
QY 61 ttactcagaatgacgttaggaagagaagtgtgtctctcttcaggcaaatgtcttctat 120
Db 218 TTATTGAGATGACITTAGGAGAGAGCAATGTCTCTCTTCAGSCAATGTCTTCTAT 277
QY 121 actgtgctggcagaaatgttttaagatgggatcttttcaccagagcaaatataaacaaga 180
Db 278 ACTGTGATGCGCAGAAATGTTTAAAGATGGGATCTTTCACAGAGCAAAATTAAGAAGA 337
QY 181 actgagagactcatgtgcagaccacaaagggtgacgatgctgttggaatgctcgatt 240
Db 338 ACTGAGGAGCTGATTCGGCAGACCAACAGGTGTATGATGATTTGGATGCTTGACATT 397
QY 241 gaggaagtaacttcacgagaaactgtctgcaggcactgcgagatgtagaagtaaatatata 300
Db 398 GAAGAGGTAACTTATCAGAACTGTTTTCAGGCACCTGCGAGATGTGGAAGTATATA 457
QY 301 gtggaaggaacatgttagactttcccccagcatgtatctctgacaaaagaagtagcagca 360
Db 458 GTGGAACGAACCATGCTAGACTTTCCCGCAGCATGTCTCTCTGACAAAGAACGACGCA 517
QY 361 gcaatcacagaacagacaaaagactttctgtttgattgagatgagcatgagcagaga 420
Db 518 GCNAGTACAGAACGACACAAAGACATTTCTGTTTGTATGATGAGATGAGATGAGAGAA 577
QY 421 gatattttgagaattgttctatttacaggaaacctgtgatctggggagagataaaacct 480
Db 578 GATATCTTCTGAGAATTGCTCGTTTAAAGGACACCTGTGATTGGGGAAGATCAAGCCT 637
QY 481 gagggcagacgatacttggaaaagtcaattaaaatggggaagaaatgggctccatctt 540
Db 638 GAAGCCCAAGATACTTAGAAAAATCAGTTAAATGGGAAAAAGGAATGGGCTCCATCTT 697
QY 541 cctgaacaagtacagaatgaatcaaatcaatgaagaaaaagaatgagtgagctatgatt 600
Db 698 CCGGAACAAAGTACAGAAATGAATCAAGCAATGAAGAAAGAAATGAGTGTGATGATT 757
QY 601 gattttaacaaaaacctcaatgaggatgataccttctgtattttccaaggtgaactt 660
Db 758 GACTTCAACAAAAACCTCAATGAGGACGATACCTTCTCTGTATTATTTCCAGGCTGACTT 817
QY 661 ggtgctcttctgatgattttcattgacagtttagaaaagacagatgatgacaagtataaa 720
Db 818 GGTGCTCTCTCTGATGATTTCATTGACAGTTTAGAAAAGACACATGACAAACAAGTATAA 877
QY 721 attacctataaatatcacactatttccctgtcatgaagaaatgttgtatccctgaaccc 780
Db 878 ATTACCTTAAATATCACACTATTTTCTGTCTCATGAAGAAATGTTGTATCCCTGAAACC 937
QY 781 agaagaagatgaatggcgttttaatacaagtgcaagagaaaaaacacataatttg 840
Db 938 AGAAGGAAATGGAATGGCTTTTAAACAGGTGCAAGAGGAAACACCATAAATCTCTG 997
QY 841 cagcagctactcccactgcgaaccaaaggtggccaaactactcgtgttatagcacacatgct 900

```

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: January 14, 2002, 18:59:51 ; Search time 3021.82 Seconds  
(without alignments)  
11546.530 Million cell updates/sec

Title: US-09-833-782-1  
Perfect score: 2115  
Sequence: 1 atgacgccggcgccctttt.....gagccctgcactctcgtga 2115

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vl.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_fun.\*
- 18: em\_in.\*
- 19: em\_om.\*
- 20: em\_or.\*
- 21: em\_ov.\*
- 22: em\_pat.\*
- 23: em\_ph.\*
- 24: em\_pl.\*
- 25: em\_ro.\*
- 26: em\_sts.\*
- 27: em\_sy.\*
- 28: em\_un.\*
- 29: em\_vl.\*
- 30: em\_htgo\_hum.\*
- 31: em\_htgo\_inv.\*
- 32: em\_htgo\_rod.\*
- 33: em\_htg\_hum.\*
- 34: em\_htg\_inv.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2114.2	100.0	2890	9	HSA300837	AJ300837 Homo sapi
2	2109.4	99.7	2864	6	AXI06913	AXI06913 Sequence
3	1858.2	87.9	2732	4	AB000170	AB000170 Porcine m
4	1858.2	87.9	3819	4	PIGSABP	D11336 Sus scrofa
5	1829.8	86.5	2810	4	AB000172	AB000172 Porcine m
6	1828.4	86.4	2709	4	AB000171	AB000171 Porcine m
7	1828.4	86.4	2862	4	AB000173	AB000173 Porcine m
8	1828.2	86.4	2604	4	AB000174	AB000174 Porcine m
9	1826.8	86.4	2728	4	AB000175	AB000175 Porcine m
10	1799	85.1	3507	4	RABEPD	D13310 Rabbit mrna
11	1653.4	78.2	2448	10	RNRNAME	X87157 R.norvegicu
12	1526	72.2	2637	9	BC004985	BC004985 Homo sapi
13	1526	72.2	2646	9	BC001844	BC001844 Homo sapi
14	1344.2	63.6	5343	9	AB033052	AB033052 Homo sapi
15	1178.6	55.7	1658	9	AK027602	AK027602 Homo sapi
16	857.6	40.5	2850	5	AB030904	AB030904 Xenopus l
17	800.4	37.8	2490	4	PIGTHOL	D21871 Pig mrna fo
18	790.8	37.4	2070	6	I12383	I12383 Sequence 6
19	790.8	37.4	2462	9	HSU29366	U29366 Human Chine
20	790.8	37.4	2515	9	HSTHIMOPS	250115 H.sapiens m
21	790.8	37.4	2535	9	BC000135	BC000135 Homo sapi
22	790.8	37.4	2577	9	BC002391	BC002391 Homo sapi
23	779.6	36.9	2419	10	AF314187	AF314187 Mus muscu
24	759.2	35.9	2314	10	RATMEP	M61142 Rat metallo
25	694.6	32.8	2122	9	BC000583	BC000583 Homo sapi
26	372.8	17.6	138295	2	AC024448	AC024448 Homo sapi
27	372.8	17.6	167656	9	AC079468	AC079468 Homo sapi
28	372.8	17.6	179969	9	AC016643	AC016643 Homo sapi
29	372.8	17.6	196898	2	AC010357	AC010357 Homo sapi
30	321	15.2	391	4	AB00041S10	AB000420 Porcine D
31	263.8	12.5	109525	9	AC008958	AC008958 Homo sapi
32	263.8	12.5	196898	2	AC010357	AC010357 Homo sapi
33	226.4	10.7	284	4	AB00041S04	AB000414 Porcine D
34	180.4	8.5	138295	2	AC024448	AC024448 Homo sapi
35	179.4	8.5	226	4	AB00041S11	AB000421 Porcine D
36	172.2	8.1	204	4	AB00041S12	AB000422 Porcine D
37	153.6	7.3	3815	8	SCPRD1	X76504 S.cerevisia
38	153.6	7.3	316613	8	SCCHRII	X59720 S.cerevisia
39	152.2	7.2	185	4	AB00041S08	AB000418 Porcine D
40	149.6	7.1	385	4	AB00042S08	AB000433 Porcine D
41	139.4	6.6	173	4	AB00041S05	AB000415 Porcine D
42	133.2	6.3	112022	9	HSAJ9611	AJ009611 Homo sapi
43	133.2	6.3	177540	9	AC006538	AC006538 Homo sapi
44	132.4	6.3	201320	2	AC073816	AC073816 Mus muscu
45	132.4	6.3	247196	2	AC073822	AC073822 Mus muscu

ALIGNMENTS

RESULT	1					
HSA300837		HSA300837	2890 bp	mrna	PRI	26-JAN-2001
LOCUS		Homo sapiens mrna for neurolysin.				
DEFINITION		AJ300837				
ACCESSION		AJ300837.1	GI:12583611			
VERSION		neurolysin.				
KEYWORDS		human.				
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
AUTHORS		1 (bases 1 to 2890)				
TITLE		Chen, J.M., Rawlings, N.D. and Barrett, A.J.				
JOURNAL		Cloning and sequencing of human neurolysin, an oligopeptidase of				
REFERENCE		family M3				
AUTHORS		Unpublished				
TITLE		2 (bases 1 to 2890)				
JOURNAL		Chen, J.M.				
TITLE		Direct Submission				
JOURNAL		Submitted (23-JAN-2001)-Chen J-M., MRC-Molecular-Enzymology				

Laboratory, The Babraham Institute, Babraham, Cambridge, CB2 4AT,  
UNITED KINGDOM

FEATURES  
source  
1. .2890  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
126. .2240  
/EC\_number="3.4.24.16"  
/function="oligopeptidase"  
/codon\_start=1  
/evidence="experimental"  
/product="neurolysin"  
/protein\_id="CAC27329.1"  
/db\_xref="GI:12583612"  
/translation="MIARCLLAVSLRRVSGRILLRLMTLGRVMSPLQAMSSVTVAG  
RNVLRWDLSPQIKTRTEELIVQTKYQVDAVGLGIEEYVYENCLOLADVEKYIIVE  
RTMLDFQHVSSQKVEAASTEADKRLSRFEDIEMSMRGIDIFERIVHQLQEDTGLKIKP  
EARYLEKSKMGRNGLHPEQVONEIKSMKRMSLCLDFNKNLNMEDDTFLVFSKA  
ELGALPDFIDSLKEDDDDKYKITLKYPHYFPVMKKCCIPETRRMEMAFNTRKEEN  
TIILOQLPRTKVKALGLYSTHADVLENNATKSTSRVAFDLDDLKSLKPLGEAER  
EFILNLAKKCKRGRFVDCGKINAWDLIYYMTOTEEKLYSIDOEFLKEYPEIEVVTG  
LLNTYQELLGLSPQMTDAHWKNSVTLIYFKDKATGEVLQGFYLDLIPREGYNHAA  
CFGLQPCLLPDSGRMAVALVYVNSQPVAGRPSLRKHDEVRTYFHEFGVHMQICA  
QTFARFSGNTNEDFEVPSQMLNWNVDVSLRLSKHYKDGSPHADLLEKLVAS  
RLVNTGLTLRLQIVLSKVDQSLHTNTSLDAASEYKCYSEILGVAATPGTNMPATFGH  
LAGGYDQGYLYLWSEVFSMDMFYSCFKKEGIMNPEVMKYRNLLIKPGGSLDGMML  
HNFLKRPNQKAFIMSLGLHAP"

BASE COUNT 861 a 571 c 662 g 796 t

ORIGIN

Query Match 100.0%; Score 2114.2; DB 9; Length 2890;  
Best Local Similarity 99.98; Pred. No. 0;  
Matches 2113; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgctgcgcgcgttccttggctgtgcgaagcctccgcagagttgggtccaggatt 60  
Db 126 ATGATCGCGCGTGCCTTTGGCTGTGCGAAGCCTCCGCAGAGTTGGTTCAGGATT 185

Qy 61 ttactcagaatgagcttaggaagaagatgatgtctctctctctcaggaatgtcttcctat 120  
Db 186 TTACTCAGAATGAGCTTAGGAAGAAGATGATGCTCTCTCTTCAGGCAATGCTCTTCCTAT 245

Qy 121 actgtggctgcagaaatgttttaagatggatctctccaccagagcaaatataaacaga 180  
Db 246 ACTGTGGCTGCAGAAATGTTTTAGATGGATCTTTCACCAGAGCAAAATTAACACAGA 305

Qy 181 actgaggagctcattgtgcagaccaaacaggtgtacgatgtgttggaaatgctcgtgatt 240  
Db 306 ACTGAGGAGCTCAITGTGCAGACCAACACAGGTGTACGATGCTGTGGAACTGCTCGTATT 365

Qy 241 gagaaagttaacttcagagaactgtctgcaggaactgtgcagatgtagaagtataatata 300  
Db 366 GAGAAAGTAACTTACGAGAACTGTCTGCAGCACTGGCAGAGTGTAGAAGTAAAGTATATA 425

Qy 301 gtgaaagaccatctagacttcccagcatgtatctctctcagaaagaagatcagagca 360  
Db 426 GTGAAAGACCATCTAGACTTTCCCGAGCATGTATCCTCTGACAAAGAAGATACGAGCA 485

Qy 361 gcaagtcagagcagacaaagacttctctgttttgatatgtgagatgagcatgagagga 420  
Db 486 GCAAGTCAGAGCAGACAAAAGACTTCTCGTTTTGTATATTGAGATGAGCATGAGAGGA 545

Qy 421 gatattattgagaattgttcaattacagaaacacctgtgtctctgggaagataaaacct 480  
Db 546 GATATATTGAGAAATGTTTCATTTACAGAAACCTGTGATCTGGGAAGATAAAACCT 605

Qy 481 gagccagacatcttgaaaagtcaattaaatggggaagaaatgggctccatctt 540  
Db 606 CAGCCAGACATCTTGGAAAGATCAATTAATGGGAAAGAAATGGGCTCCATCTT 665

Qy 541 cctgaacaggtacagaatgaaatcaaatcaatgaagaaagaatgagtgagctatgtatt 600

Db 666 CCTGAACAAGTACAGAATGAAATCAATGAAGAAAGAATGAGTGAAGTATGATT 725  
Qy 601 gattttaaaaaaacctcaatgagatgatcccttctctgtatttccaaagctgaacct 660  
Db 726 GATTTTAAACAAAACCTCAATGAGGATGATACCTCTCTGTATTTCACAGGCTGAACCT 785  
Qy 661 ggtgctctctctgatgatttcattgacagtttagaagaagacagatgatgacaaagtataaa 720  
Db 786 GGTGCTCTTCTGATGATTTCATTGACAGTTTAGAAAAGACAGATGATGACAAAGTATAAA 845  
Qy 721 attaccttaaaatatccacactatttccctgtcatgaagaagaattgttgcctcgaacc 780  
Db 846 ATTACCTTAAATATCCACACTATTTCCTGTCATGAAGAAATGTTGTATCCCTGAAACC 905  
Qy 781 agaagaagatggaatggcttttaatacagaagtgcaagaaggaagaaacaccataatttg 840  
Db 906 AGAAGAAGATGGAATGGCTTTTAAACAAGGTGCAAGAGGAAACACCATTAATTTTG 965  
Qy 841 cagcagctactccctcactgcgaaccaaggtggccaaactactcggttatagcacacatgct 900  
Db 966 CAGCAGCTACTCCCACTGCGAACAAGGTGCGCAAACTACTCGTTATAGCACACATGCT 1025  
Qy 901 gactctgctctgaaatgaacactgcgaagcgaagcgcgcgttaacagcytttctatagat 960  
Db 1026 GACTTCGCTCTTGAATGAACACTGCAAGAGAGCACAAAGCCGCTAACAGCCTTTCATAG 1085  
Qy 961 gatttaagcagaagtttaaaaccccttggtgaaagcagaagagatttatttgaatttg 1020  
Db 1086 GATTTAAGCCAGAGTTTAAACCCCTTGGGTGAAGCAGAGAGAGTTTATTTTGAATTTG 1145  
Qy 1021 aagaaaaaggaatgcaaaagacaggggttttgaatatgatgggaaatacaatgcctgggat 1080  
Db 1146 AAGAAAAGGAATGCAAGACAGGGGTTTTTCAATATGATGGGAAATCAATGCTGGGAT 1205  
Qy 1081 ctatattactacatgactcagacagagagaactcaagtaattccatagacaaagagttctctc 1140  
Db 1206 CTATATTACTACATGACTCAGACAGAGAACTCAAGTATTCCATAGACCAAGAGTTCTCTC 1265  
Qy 1141 aagaataacttcccaattgaggtgctcactgaagcttgcgtgaacacaccacagagattg 1200  
Db 1266 AAGAAATACTTCCCAATTGAGTGGTCACTGAAGCTTGTGTAACACCTCACAGAGATTG 1325  
Qy 1201 ttggagcttcttcatgtgaacaaatgacagatgctcatgtttggaaacagagttgtacacct 1260  
Db 1326 TTGGAGCTTTCATTTGAACAAATGACAGATGCTCATGTTTGAACAAAGAGTTTACACTT 1385  
Qy 1261 tatactgtgaaggataaagctacaggaagaatattgggaacagttctatttggacctctat 1320  
Db 1386 TATACTGTGAAGGATAAAGCTACAGGAGAAAGTATTGGACAGAGTTCTATTTTGGACCTCTAT 1445  
Qy 1321 ccaaggaagaagaataacaatcatcgccctgtctcgtgtctcagcctggtgctctctg 1380  
Db 1446 CCAAGGAAGAAGAAATACAATCATCGCGCTGCTTCGGTCTCCAGCCTGCGCTTCTG 1505  
Qy 1381 cctgatgaagcccgatgatgagcagtgctcctcctcgtgtgaactctcagaccagtg 1440  
Db 1506 CCTGATGGAAGCCCGATGATGGCAGTGGCTGCCCTCGTGGTGAACCTTCACAGCCAGTG 1565  
Qy 1441 gcaggtcgtcctctctctctgagacagcagaggtgaggaacttacttctcatgagttgggt 1500  
Db 1566 GCAGGTCTCTCTCTCTCTGAGACACAGCAGGAGTGAGGAGCTTACTTTTCAATGAGTTGGT 1625  
Qy 1501 cactgtatgatacagatttgcacagactgatttgcacagatttgcacagatttgcacagatttgc 1560  
Db 1626 CACTGTATGATCAGATTGTGACAGACTGATTTTGCACGATTTAGCGGAACAAATTTG 1685  
Qy 1561 gaaactgactttgtagaggtgcccactcgaatgcttgaataattgggtgagacgtcgat 1620  
Db 1686 GAAACTGACTTTGTAGAGGTGCCATCGCAATGCTTGAATAATTGGGTGGGACGTGAT 1745  
Qy 1621 tccctcgaagattgtcaaaacattataagatggaagccctattgcagacagatctgctt 1680

Db 1746 TCCTCCGAAGATTGCAAAACATTTATAAGATGAAGCCCTATTTCACAGACCATCTGCTT 1805

Qy 1681 gaaaaactgttcttcttaggtgttaacacaggtcttcttgcacctgcgcagattgtt 1740

Db 1806 GAAAAACTTGTCTTCTAGGCTGTCAACACAGCTCTTCTGACCTGCGCCAGATTGTT 1865

Qy 1741 ttgagcaaatgttgatcagctcttctataccacacacatgcgtgagatgcgcaagtgaat 1800

Db 1866 TTGAGCAAAAGTTGATCAGCTCTTATACCAACACATCGCTGGATGCTGCAAGTGAATAT 1925

Qy 1801 gcaaatctgctcagaaatattaggttgagctactctccagggcacaaatgacagct 1860

Db 1926 GCCAATATCTGCTCAGAAATATTAGAGTTGACAGTCTCCAGGCACAAATATGCCAGT 1985

Qy 1861 accttggacatttggcaggggtatcagtggtgccaatattatggatatcttggagtga 1920

Db 1986 ACCTTTGGACATTTGGCAGGGGATACGATGTCGCAATATTATGGATATCTTTGGAGTGA 2045

Qy 1921 gttatttccatgatatgttttacagctgttttaaaaaaagaggaatgaatcaagag 1980

Db 2046 GTATTTTCCATGATGATGTTTACAGCTGTTTAAAAAGAGGGATATATGAATCCAGAG 2105

Qy 1981 gttggatgaaacacagaaaccttaacctgaacccctggggggtctctgacgcatggac 2040

Db 2106 GTTGGATGAATACAGAACTAATCTCTGAAACCTGGGGGATCTCTGACGCGATGGAC 2165

Qy 2041 atgtccacaatttctgaaacgttgccaaacccaaagcgttcttaagtgaagagc 2100

Db 2166 ATGCTCCCAATTTCTTGAACGTGACGCAAAACCAAAAGCCGTTCTTAATGATGAGTAGGC 2225

Qy 2101 ctgcatgctcgtga 2115

Db 2226 CTGCATGCTCCGGA 2240

RESULT 2

LOCUS AX106913 2864 bp DNA PAT 30-APR-2001

DEFINITION Sequence 2 from Patent WO0123590.

ACCESSION AX106913

VERSION AX106913.1 GI:13922466

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2864)

AUTHORS Kapeller-Libermann, R., White, D. and Silos-Santiago, I.

TITLE 22196, a novel human aminopeptidase

JOURNAL Patent: WO 0123590-A 2 05-APR-2001

FEATURES

source Location/Qualifiers

1..2864

/organism="Homo sapiens"

/db\_xref="taxon:9606"

300..2414

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAC37750.1"

/db\_xref="GI:13922467"

/translation="MIARCLLAIVSLRRVGGSRILLRMTLGRVMSPLQAMSSYTVAG RNVLWDLSPQIKTTELEIVQTKQVDAVGLMGEVITYENCLOALADVEKVIYE RTMLDPQIVSSDKRNLSTADKRLSRFDIEMSRGDI FERI VHLQETDLGKIKP EARYLEKISIRKQKRLHLPQVQNEIKSMKRMSELIDFNKLNLEDITL VFSKA ELGALPDDTIDSLKDDDKYKTLTKYPHYFVPMKCCIPETRRRMEAFNTRKEEN TILLOLLPLRTKVALKLGYSHADPVLPMNTAK STSRVTAFLDDL SOKLPLGEAER EFILNKKCEKDRGFEYDGIKINAWDLYYMTQTELYSIDQELKEVEPLEVYTEG LNTYQELLGLLSPQMDAHWNKSVLTITVTKDKATGEVIGQFYLDLYPREGKYNHAA CFTGLQGLLSPGSRMVAALVFNFSQPVAGRSLLRHDEVRTYFPEFGHWHQICA QTFDFAFSGTNETDFEVSQMLNENWVDLSRLSKYKDGSP IADDLLEKLVAS RLAVNTGLLRLQIVLSKVQDSLSLNTSLDAASEYAKYCSILGLVATPGTNMPATFGH LAGYDQGVYGVYLSVFSVMDMFYSCFKKEGIMNPVGVMKRYNLILKPGSLDGMDL HNFLKREPQKAFILMSRGLHAP"

CDS

BASE COUNT 854 a 556 c 693 g 761 t

ORIGIN

Query Match 99.7%; Score 2109.4; DB 6; Length 2864;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 2110; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 atgatgcgccgggtgctcttggctgtgcgaagcctccgcagagttgggttccaggatt 60

Db 300 ATGATGCGCCGGTGCCTTTTGGCTGTGCGAAGCCCTCCGAGAGTTGGTGTTCAGGATT 359

Qy 61 ttactcagaatcacgttaggaagagaagtgtctctctcctccagggcaaatgtctcctat 120

Db 360 TTACTCAGNATCAGCTTAGGAAGAGAAAGTGTCTCTCTTCAGGCAATGCTTCCTTAT 419

Qy 121 actgtggctggcagaaatgttttaagatgggatcttccaccagagcaaaattaaacaaga 180

Db 420 ACCGTGGCTGGCAGAAATGTTTAAAGATGGGATCTTTCCGACAGAGCAAAATTAACAAGA 479

Qy 181 actgaggagctcattgtgcagaccacaaacaggtgtcagatgctgttggaaatgctcgatt 240

Db 480 ACTGAGGAGCTCATTGTGCAGACCAGAGGTGTACGATGCTGTGGAAATGCTCGGATT 539

Qy 241 gaggaagtaacttacagagaactgtctgcaggcactgcgcagatgtagaagtataata 300

Db 540 GAGGAAGTAACCTTACGAGAACTGTCTGCAGGCACCTGGCAGATGTAGAACTAAAGTATATA 599

Qy 301 gtggaaaggaccatgctagacttcccacagcatgtatcctctgacaaagaagtaagagca 360

Db 600 GTGGAAGGACCATGCTAGACTTTCCCGACATGTATCTCTGACAAAGAGTACGAGCA 659

Qy 361 gcaagtacagaagcagacaaaagacttctctgttttgatattgagatgagcatgaagca 420

Db 660 GCAAGTACAGAAGACAGACAAAAGACTTTCTCGTTTTGTATTGAGATGAGCATGAGAGGA 719

Qy 421 gatatttgagagaattgttcattcacaggaacctgtgactctgggggaagataaaacct 480

Db 720 GATATATTGTAGAGAAATGCTTCATTTACAGGAACCTGTGTCTGGGGAAGATAAAACCT 779

Qy 481 gaggccagacgatacttggaaagtcaatataaatggggaaagaagaatgggtccatctt 540

Db 780 GAGGCCAGACGATCTTGGAAAGTCAATTAATAATGGGAAAGAAATGGGTCCATCTT 839

Qy 541 cctgaacaagtcagacaatgaatacaaatcaatgaagaagaagaatgagtgatctattt 600

Db 840 CCTGAACAAGTACAGAAATGAAATCAAAATCAATGAAGAAAGAAATGAGTGTATGTAAT 899

Qy 601 gattttaacaaaaacctcaatgagatgatcaccttctctgtattttccaggcgtgaactt 660

Db 900 GATTTTACAAAACCTCAATGAGGATGATACCTCTCTGTATTTCACAGGCTGAACCT 959

Qy 661 ggtgctcttctgagtgatttctcattgacagtttagaagaagacagatgatgacagataaa 720

Db 960 GGTGCTCTTCTCTGATGATTTCAATGACAGTTTAGAAAAGACAGATGATGACAAAGTATAA 1019

Qy 721 attacctaaaaataaccacactatttccctgtcattgaagaagaatgttgatccctgaacc 780

Db 1020 ATTTACCTTAAATATCCACACTATTTCCCTGTCTATGAAGAAATGTTGTATCCCTGAAC 1079

Qy 781 agaagaagatgaaatggcttttaatacaaggtgcaagaggaagaaacacataatttg 840

Db 1080 AGAAGAAGATGGAATGCTTTTAATACAGGTGCAAGAGGAAACACCAATATTTTG 1139

Qy 841 cagcagctactccactgcgaacaaagggtggccaaactactcgtgttatagcacacatgct 900

Db 1140 CACAGCTACTCCCACTGCGAACCAAGTGGCCAAACTACTCTGCTTATAGCACACATGCT 1199

Qy 901 gacttgccttgaatgaacactgacaaagcagacagccgcgttaacagcgtttcttagat 960

Db 1200 GACTTCTCTCTGAAATGAACACTGCAAGAGACAGCCGCGCTAACAGCTTCTTAGAT 1259

Qy 961 gatttaagccagaagttaaaaccttgggtggaagcagacagagatttttgaatttg 1020

Db 1260 GATTTAAGCAGAAAGTAAACACCTTGGGTGAAGCAGAGAGTATTTTGAATTG 1319  
QY 1021 aagaaagaaatgcaaacagaggttttgaatgatgagaaatcaatcctggat 1080  
Db 1320 AAAAAAAGAAATGCAAGACAGGGGTTTTGAATATGATGGGAAATCAATGCTGGAT 1379  
QY 1081 ctatattactactgactcagacagaggaactcaagtattccatagacacaaagttccctc 1140  
Db 1380 CTATATTACTACTGACTCAGACAGAGGAACCAAGTATTCATAGACCAAGAGTTCCTC 1439  
QY 1141 aagaaatacttcccaattgaggtgctcaatgaaggttctgtaacacacctaccaggagttg 1200  
Db 1440 AAGAAATACTTCCCAATGAGGTGGTCACTGAAGGCTTCTGTAACACCTACCAGGAGTTG 1499  
QY 1201 ttggagcttcttgaacaaatgacagatgctcatgtttgaaacaaagatttacaatt 1260  
Db 1500 TTGGGACTTTCATTGACAAATGACAGATGCTCATGTTGGGAAAGAGTGTACACTT 1559  
QY 1261 tatactgtgaagataaaactacagaggaagtattgggacagttctatttggacctctat 1320  
Db 1560 TATACGTGTAAGGATAAAGCTACAGGAGAGTATTTGGGACAGTCTATTTGGACCTCTAT 1619  
QY 1321 cbaaggaagaaaaatacaatcaatcggtctgttgggttctccagcctggtgcttctg 1380  
Db 1620 CCAAGGGAAGGAAAAATACAAATCATGCGGCTGCTTGGTCTCCAGCTGCTGCTTCTG 1679  
QY 1381 cctgatgaagcagcagatgagcagtgctgctgctcctgctgctgctgctgctgctgctg 1440  
Db 1680 CTTGATGGAAGCGGATGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1739  
QY 1441 gagggtcgctcctctctctgagacacagcagaggtgaggaacttcttcatgagttggt 1500  
Db 1740 GCAGGTGCTCCCTCTCTCTGAGACAGCAGAGGTGAGGACTTACTTTCATGAGTTGGT 1799  
QY 1501 cactgatcatcagatttgcacagactgattttgacgactattttgacggaacaaatgtg 1560  
Db 1800 CACGTGATGCATCAGATTGTGCACAGACTGATTTTGCACGATTTAGCGGAACAAATGTG 1859  
QY 1561 gaaactgactttagaggtccatcgcaaatgctgtaaatggtggtggtgagacatgat 1620  
Db 1860 GAACATGACTTTGTAGAGTGCATCGCAATGCTGTAATTTGGGTGGTGGGACGTCGAT 1919  
QY 1621 tccctcgaagattgtcaaaacattataaagatggaagccctatttcgacagatctgctt 1680  
Db 1920 TCCCTCCGAAGATTGTCAAAACATTATAAAGATGGAAGCCCTATTTCACAGCATCTGCTT 1979  
QY 1681 gaaaaacttgttcttaggtgctgcaacacaggtcttctgacccctgcgcagattggtt 1740  
Db 1980 GAAAAACTTGTGCTTCTAGGCTGCTCAACACAGCTCTTCTGACCTCGCCAGATTGTT 2039  
QY 1741 ttgagcaaatgtgatcagctctcttaccacacacatcgctgagtgctgcaagtgaatat 1800  
Db 2040 TTGACCAAGTTGATCAGTCTCTTCAACCAACATCGCTGGATGCTGCAAGTGAATAT 2099  
QY 1801 gcaaaactgctcagaatatttagagttgagctactccaggcacaataatgcagact 1860  
Db 2100 GCAAAATCTGCTCAAAATATTAGAGTTGACAGTCTTCCAGGCACAAATATGCCACT 2159  
QY 1861 accttggacatttggcaggggatacatgaggaataattatgatacttcttggagtgaa 1920  
Db 2160 ACCTTTGGACATTTGGCAGGGGGATACCATGGCCAAATATTATGGATATCTTTGGAGTGA 2219  
QY 1921 gtatttccatgatatgttttaccagctgttttaaaaaagaggagataatgaatccagag 1980  
Db 2220 GTATTTTCCATGGATATGTTTACAGCTGTTTAAAAAAGAGGATTAATGAATCCAGAG 2279  
QY 1981 gttggaatgaatacagaaactaatcctgaaacctggggatctctctgagcggcatggac 2040  
Db 2280 GTTGAATGAATACAGAAACCTAATCTGAAACCTGGGGATCTCTGCACGGCATGGAC 2339  
QY 2041 atgctcccaatttctgaaacgtgagcacaacaaacaaagcttccctaattagtagagc 2100

Db 2340 ATGTCCCAATTTCTGAAACGTGAGCCAAACCAAAAGCGTTCTCTATGATGAGGC 2399  
QY 2101 ctgcatgctcgtga 2115  
Db 2400 CTGCATGCTCCCGTGA 2414  
RESULT 3  
AB000170 2732 bp mRNA MAM 05-FEB-1999  
LOCUS Porcine mRNA for endopeptidase 24.16, complete cds.  
DEFINITION AB000170.1 GI:1783121  
ACCESSION AB000170  
VERSION endopeptidase 24.16 type M3; endopeptidase 24.16 type M1.  
KEYWORDS Sus scrofa Adult Liver cDNA to mRNA.  
SOURCE Sus scrofa  
ORGANISM Sus scrofa  
REFERENCE 1 (bases 1 to 2732)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
Hirose, S.  
Direct Submission  
Submitted (27-DEC-1996) to the DDBJ/EMBL/GenBank databases.  
Shigenisa Hirose, Tokyo Institute of Technology, Department of  
Biological Sciences, 4259 Nagatsuta-cho, Midori-ku, Yokohama,  
Kanagawa 226-8501, Japan (E-mail:shirose@bio.titech.ac.jp,  
Tel:045-924-5726, Fax:045-924-5824)  
2 (sites)  
Kato, A., Sugliura, N., Saruta, Y., Hosoi, T., Yasue, H. and Hirose, S.  
Targeting of endopeptidase 24.16 to different subcellular  
compartments by alternative promoter usage  
J. Biol. Chem. 272 (24), 15313-15322 (1997)  
97326108  
FEATURES  
Location/Qualifiers  
1..2732  
/organism="Sus scrofa"  
/db\_xref="taxon:9823"  
/dev\_stage="Adult"  
/tissue\_type="Liver"  
1..2732  
/note="corresponding to exon1.5-16 of this gene;  
endopeptidase 24.16 type 1"  
175..2289  
/standard\_name="endopeptidase 24.16"  
/note="oligopeptidase M: neurolysin :sBAP(soluble  
angiotensin-binding protein) :MEP(microsomal  
metalloendopeptidase)"  
/codon\_start=1  
/product="endopeptidase 24.16 type M1"  
/protein\_id="BAA19060.1"  
/db\_xref="GI:1783123"  
/translation="MIVRCISAAARRLHVRGSGTLLRMTLGRAMSPLOAMSSVTVDG  
RNVLRDLSPEQIKRTEELIAQTKQVDDIGMDLIEVTVENCLOADEVKYIVE  
RTMLDPQHVSSKEVRASTEADKRLSREDIESMRDELFLRLVRLKTCDDKINP  
EARYLEKSVKMGKRLNGHLPEQVQNEIKAMKRMSELCDFNKNLNEDDTFLVFSKA  
ELGALPDFIDSLKTDNNKTKITLKYPHYVPMKCCIPETRRKMEAFNTRKEEN  
TIILQELLPLRAKVALIGYSTHADFLYFVPMKCCIPETRRKMEAFNTRKEEN  
ETILNLKKECEKGEYDQKINAWDLHYMTQTEELKYSVDQELKYEPIEVWTEG  
LLNIYQELLGLSFEQVTDHVNKSVTLTYTKDKATGEVLGQFVLDLYPREKYNHAA  
CFGLQGLCLPDGSRMSVAALVYVNFQPRAGRPSLLRHDEVITYEFHFGVHQAICA  
QTDFAFPSTNVETDFVPSQMLENNWDTLSRLSKHYKDGSPITDLDLEKLVAS  
RLVNTGLLRLQIVLSKVQSLHTNTSLDAASEYAKYCTEILGVAATPGTNMPATFGH  
LAGYQDQYGYLWSEVFSMDPVSFCFKKEGINPEVGMKRYNRLILKPGGSLDGMML  
QNFLKRPNQKAFMLSRGLHAP"  
transit\_peptide 175..285  
CDS 244..2289  
/standard\_name="endopeptidase 24.16"  
/note="alternative product for this mRNA; oligopeptidase M  
:neurolysin :sBAP(soluble angiotensin-binding protein)  
:MEP(microsomal metalloendopeptidase)"  
/codon\_start=1  
/product="endopeptidase 24.16 type M3"  
/protein\_id="BAA19061.1"  
/db\_xref="GI:1783123"



/translation="MTLGREAMPLOAMSSVTVDRNVLNRWDLSPQEIQRRTTEELIAQ  
TKQVDDILMDIERVENCLOALADVEVYIIVERTMLDFPQHVSKMSKRNGLHLPEQ  
DKLRSRFDIEMSRDIFLRLVKETCDLGIKIPKPEARYLEKHQVSMKRNGLHLPEQ  
VQNETKAMKRMSELCDIDFNKLNEDDTFLVFSKALGALPDFDIDSEKIDDDNKYKI  
TLKPHYFPVMKGCIPETRRMENAFTNRCKEENTIILOELLPIURAKVALLGYSTH  
ADFLVEMTAKSTHHVTAFLDLSOKLPLGEAEREFLINLLKKKEKFEYDGKIN  
AWDLHYMTQTEELKYSQVQELILKAYFEVTEGLNIIYQELLLGSFSEQVTDHVMN  
KSVLTITVKDRATGVLQFIDLAYPRGKYNHAAFCGLQPCLLPDGSRMMSVAALV  
VNFSPRAGRSLRLRHDSVRTYFHEFGHVMQICAOITDFARESGFNVTEDVEVPSOM  
LENMWDPSLRLSKHYKDGSPITDDLLLEKIVASRLVNTGLLIRQIVLVSQVDSLH  
TNTSLDAASEYAKVCTELGVAATPGTNPATFGLAGGYDGQYITGYIWSVEFSDMF  
YSCRFKEGIMPEVGMKYNRLNLLPGSLDGMQLNFKLRPNOKAFIMRGLHAP"  
BASE COUNT 826 a 545 c 641 g 720 t  
ORIGIN

Query Match 87.9%; Score 1858.2; DB 4; Length 2732;  
Best Local Similarity 92.3%; Pred. No. 0;  
Matches 1953; Conservative 2; Mismatches 160; Indels 0; Gaps 0;  
QY 1 atgatccgcggtgcttcttgctgtgcgaagccctccgcagagttggtgtccaggatt 60  
DB 175 ATGATGTCGGTGGCTTCGGCTGGAGAGCGCTCCACAGGGTGGTGGTCCGGGATT 234  
QY 61 ttactcagaatgacgttaggaagaggaagtgtgtctctctccaggaatgtcttccat 120  
DB 235 TTATTGAGATGACITTAGGAGAGAGCAATGTCTCTCTTCAGGCATGTCTTCTAT 294  
QY 121 actgtgctgagcaaatgtttaaagatggatcttccaccagagcaaaataaaacaaga 180  
DB 295 ACTGTGATGGCAGAAATGTTTTAAGATGGATCTTTCACCAGAGCAAAATTAAGAAGA 354  
QY 181 actgaggaactcattgtgcagaccacaaagggtgacagatgctgttggaatgtcgtatt 240  
DB 355 ACTGAGGACCTGATTGCGCAGACCAACAAAGGTGTATGATATTGGAAATGCTTGACAT 414  
QY 241 gagggaagtcaactacagagactgtctgcaggcactgcgcagatgagagataaagtata 300  
DB 415 GAAGAGTAACATTATGAGAACACTGTTTGAGGACCTGGCAGATGTGGAATGAGATATA 474  
QY 301 gtggaagagaccatgtagactttccaccagcatgtatcctctgacaaagaagtacagca 360  
DB 475 GTGGAACGAACCATGCTAGACTTTCCCGAGCATGTCTCTCTGACAAAGAAGTACGAGCA 534  
QY 361 gcaagtacagaagcagacaaaagactttctgttttgatattgagatgagatgagagga 420  
DB 535 GCAAGTACAGAGACGACAAAAGACTTTCTGCTTTTGTATTGATGATGATGAGAGAA 594  
QY 421 gatatattgagagaattgttcatttacaggaaacctgtgactctgaggggaagataaacct 480  
DB 595 GATATCTTTCTGAGATTTCTGTTTAAAGGAGACCTGTGATTTGGGGAAGATCAAGCCT 654  
QY 481 gaggccagacgatacttggaagaagtcaataaaatggggaagaagaatgggtccatctt 540  
DB 655 GAAGCCAGAAGATACTTAGAAAAATCAGTTAAATGGGAAAAAGGAATGGGTCCCATCTT 714  
QY 541 cctggaacaagtacagaatgaaatacaatgaagaaagaagaatgagtgagctatgtatt 600  
DB 715 CCGGAACAAGTACAGAATGAAATCAAAAGCAATGAAGAAAGAAATGAGTACGATGATATT 774  
QY 601 gattttaacaaaacctcaatgagatgatactctctctgtattttccaggctgaactt 660  
DB 775 GACTTCAACAAAACCTCAATGAGGACGATACCTTCTTGTATTTTCCAAAGGCTGAACCT 834  
QY 661 ggtgtcttcttgatgatttcttgacagtttagaaaaagacagatgataagtaataaa 720  
DB 835 GGTGCTCTTCTGATGATTTTCATTGACAGTTTAGAAAAACAGATGACAAAGATATAA 894  
QY 721 attaccttaaaatataccacatttccctgctcatgaagaagaatgtgtatccctgaaacc 780  
DB 895 ATTACCTTAAATATCCACACTATTTTCTGTCATGAAGAAATGTTGTATCCTCGTGAACC 954  
QY 781 agaagaaggatgaaatggtcttttaatacaaggtgcaagaggaaaaacacccaataatttg 840

DB 955 AGAAGAAAAATGGAATGCTTTTAAATACAGGTGCAAAAGAGGAAAAACCAATAATCCTG 1014  
QY 841 caggaactactccactcgcgaaccgaaggtggccaaactactcgtgttatagcacacatgct 900  
DB 1015 CAGGAGCTACTCCCACTCGGGCCAAAAGTAGGCAAACTACTTGGGCTACAGCACACATGCT 1074  
QY 901 gacttcgtccttgaatgaactgcaagagacacagcgccgtaacagcyyttcttagat 960  
DB 1075 GACTTTGCTTGTAGATGAACACTGCGAAGAGTACACACCATGTGACAGCTTCTTAGAT 1134  
QY 961 gatttaagccagaagttaaaaccccttggtggaagcagaagagaggtttattttgaatttg 1020  
DB 1135 GATTTAAGCCAGAAATTAACCCCTTGGGTGAGGCAAGACGAGAGTTCAATTTGAATTTG 1194  
QY 1021 aagaaaaagaatgcaaaagacagagggtttttgtaataatgaggaataatcaatgctcgtggat 1080  
DB 1195 AAAAAAGAAATGTGAAGAGAAAGGTTTGAATATGATGGCAAAATCAATGCTGGGAT 1254  
QY 1081 ctatattactacatgactcagacagaggaactcaagtattccatagacacaaagagttctc 1140  
DB 1255 CTACATTACTATATGACTCAGACAGAAGAACTCAAGTACTCTGTGGACCAGAGATCTCTC 1314  
QY 1141 aaggaatacttcccaattgaggtggtcactgaagggtctgtgaacacctaccaggagttg 1200  
DB 1315 AAGGAATACTTCCCAATTGAAGTGGTCACTGAAGGCTTGTCTGAACATCTATCAGGAGTTG 1374  
QY 1201 ttgggacttctatttgaacaaatgacagatgctcatggtttggacaagagtgtaacatt 1260  
DB 1375 TTGGGACTCTCTTTGAGCAAGTGACAGATGCTCATGTTTGGAAATAAAGTGTTACCTG 1434  
QY 1261 tatactgtgaagataaaactacagagagaagtattgggacagttctatttggacctctat 1320  
DB 1435 TACACGGTGAAGGATAAAGCCACAGGAGAAGTATTAGGACAGTTCTTACTTGGACCTCTAT 1494  
QY 1321 caaagggaagaaaaatacaatcatcgctgctgtctcagctgtctccagctggtcctctct 1380  
DB 1495 CCNAGGAGAGGAAAAATACAAACCATGACAGCTGCTTGGTCTCCAGGCTGGCTGCCCTCTG 1554  
QY 1381 cctgatggaagccggaatgatgagcagtggtgctgctgctggtgaacttctccagcagctg 1440  
DB 1555 CCTGATGGGAGCGCATGATGTGTGGCTGCGCTGCTGTGAACTTCTCACAGCCACGA 1614  
QY 1441 gcaagtcgtccctctctcctgagacacagcagagtgtagaacttacttctcagagttggt 1500  
DB 1615 GCAGGTGCTGCTCTCTCTCTGAGACACGATGAGGTGAGGACTTACTTCCATGAATTTGGT 1674  
QY 1501 cactgatgcatcagatttgtgcacagctgattttgcacgattttgacggaacaaatgtg 1560  
DB 1675 CAGTCAATGCATCAGATTTGTGGCAGACTGACTTTTGCAGGATTTAGTGAACAAATGTG 1734  
QY 1561 gaaactgactttagagtgccatgcgaatgctgcaaatgctgcaaatggtggtgggagctgat 1620  
DB 1735 GAAACTGTACTTTGTAGAGGTGCCATCACAATGCTTGAAGGCTGGGTGGGACACTGAT 1794  
QY 1621 tccctccgaagattgtcaaaacattataaagatggaagccctatttgcagacgactcgtt 1680  
DB 1795 TCCTCCGAAGGCTGTCAAAACATTTAAAGATGGAAGCCCTATTACAGATATCTGCTT 1854  
QY 1681 gaaaaacttgttctcttaggtgtgcaacacaggtctctctgacctggtgcccagattggt 1740  
DB 1855 GAAAACTTGTGCTTCTAGACTGGTCAACACAGGTCTCTCTGACCTTACGCCAGATGTTT 1914  
QY 1741 ttgagcaaatgtgactgtctcttctacacacatcgctggtgactgctgcaagtaatat 1800  
DB 1915 TTGAGCAAAAGTTGATCAGTCTCTCCACACCAACATCTCGCTGGATGCTGCAAGTGAATAT 1974  
QY 1801 gccaaatactgctcagaataattaggagttgcaagtctctctgacctggtgcccagattg 1860  
DB 1975 GCCAAATCTGCACAGAAATTTTAGTGTGTGACGTACTCCAGGCACAAATATGCCAGCT 2034  
QY 1861 accttggacattgagcagggggagatcgatggccaatatattgatatctttggagtgaa 1920

```
Db 2035 ACTTTTGGACACTTGGCAGGGGATATGATGCCCAATATTATGGATATCTTTGGAGTCAA 2094
Qy 1921 gtattttccatgatatttttacagctgtttttaaaagaaggataaataatccagag 1980
Db 2095 GTCCTTTCCATGGACATGTTTATAGCTGTTTAAAGAAAGAGGATATGAATCCAGAG 2154
Qy 1981 gtggaatgaatacagaaacctaactcctgaaacctggggatcctctggagcgcatggac 2040
Db 2155 GTTGAATGAATACAGAAACCTAATCTGAACCTGGGGATCGCTGGAGCGCATGGAC 2214
Qy 2041 atgctcccaattttctgaaagctgagccaaacccaaagcgttctctaatgagtagagc 2100
Db 2215 ATGCTCCAAAATTTCTTGAACGCTGAGCCAAACCAAAAGCGTTCTCTAATGAGTAGAGC 2274
Qy 2101 ctgcatgctcgtga 2115
Db 2275 CTGCATGCTCCGTAA 2289

RESULT 4
LOCUS PIGSABP 3819 bp mRNA MAM 14-APR-2000
DEFINITION Sus scrofa mRNA for soluble angiotensin-binding protein, complete cds.
ACCESSION D11336
VERSION D11336.1 GI:217708
KEYWORDS SINE; SAPP; short interspersed repeated sequence; soluble
angiotensin-binding protein.
SOURCE Sus scrofa (sub-species:domesticus) heart, liver cdNA to mRNA,
clone_lib:lambdazap II clone:lambdabAB[1,4,s2,-L1].
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 3819)
AUTHORS Sugiyama,N., Hagiwara,H. and Hirose,S.
TITLE Molecular cloning of porcine soluble angiotensin-binding protein
JOURNAL J. Biol. Chem. 267 (25), 18067-18072 (1992)
MEDLINE 92388174
REFERENCE 2 (bases 1 to 3819)
AUTHORS Hirose,S.
TITLE Direct Submission
JOURNAL Submitted (18-MAY-1992) to the DDBJ/EMBL/GenBank databases.
Shigeisa Hirose, Tokyo Institute of Technology, Department of
Biological Sciences, Ookayama, Meguro-ku, Tokyo 152, Japan
(Tel:03-3726-1111(ex.2226), Fax:03-3729-0335)
FEATURES
source
1..3819
/organism="Sus scrofa"
/sub_species="domesticus"
/db_xref="taxon:9823"
/clone="lambdabAB[1,4,s2,-L1]"
/clone_lib="lambdazap II"
/tissue_type="heart, liver"
158..2272
/note="SAPP"
/codon_start=1
/product="soluble angiotensin-binding protein"
/db_xref="GI:217709"
CDS
variation
```

```
variation
/note="deleted from clone lambdabAB1"
2168..2169
/note="substituted in clone lambdabAB1"
2676..2681
polyA_signal 2989..2994
polyA_signal 3459..3780
repeat_unit
/note="SINE(PRE-1)"
BASE COUNT 1086 a 788 c 855 g 1090 t
ORIGIN
Query Match 87.9%; Score 1858.2; DB 4; Length 3819;
Best Local Similarity 92.3%; Pred. No. 0;
Matches 1953; Conservative 2; Mismatches 160; Indels 0; Gaps 0;
Qy 1 atgctgccccgggtgccccttttgctgtgcgaagcctcccgagagtttggtggttcaggatt 60
Db 158 ATGATCGTCGGTTCGGCTTCGAGACGCCCTCCACAGGGTGGTGGTTCGGGGATT 217
Qy 61 ttactcagaatgagcttaggaagaaagtgatgtctctctctccagcgaatgtcttcctat 120
Db 218 TTATTGACAATGACTTTAGGAGAGAGCAATGCTCTCTTCAGGCATGTCTTCCTAT 277
Qy 121 actgtggctggcagaataatgttttaagatgggatctttcaccagagcaataataaacaga 180
Db 278 ACTGTGGATGCGAGAAATGTTTAAAGATGGATCTTTCACCAGAGCAAAATTAAGAGAGA 337
Qy 181 actgaggagctcattgtgcagaccacagagtgacgatgctgttggaatgactcgtgatt 240
Db 338 ACTGAGGAGCTGATGGCGCAGACCAACAGGTGTATGATGATTATGGAAATGCTTGACATT 397
Qy 241 gaggaagttaacttacgagaactgtctgcagcactggcagatggcagatgtagaagtaaatata 300
Db 398 GAAGAGGTAACTTATGAGAATGTTTGGCAGCAGCTGGCAGATGTGGAAGTGAAGTATATA 457
Qy 301 gtggaagagcaccatgctgactttcccgagcatgtatctctgcacaaagagtagacaga 360
Db 458 GTGGAACGAACCATGCTAGACTTTCCCGAGCATGTCTCTCTGACAAAGAGTAGAGCA 517
Qy 361 gcaagtacagaagcagacacaaagacttctcgtttctgatttgcagatgagcatgagaga 420
Db 518 GCAAGTACAGAGCAGACAAAGACTTTCTCGTTTGTATTTGATGATGAGCATGAGAGAA 577
Qy 421 gatatatttgagaaattgttcatttacagaaacctgtgactctggggaagataaaacct 480
Db 578 GATATCTTTCTGAGAAATGTTCTGTTTAAAGGAGACCTCTGTGTTTGGGGAAGATCAAGCT 637
Qy 481 gaggcagacgatacttggaaaagtcaattaaaaatggggaaaaaagaaatgggctccatt 540
Db 638 GAAGCCAGAAGTACTTAGAAAAATCAGTAAAAATGGGAAAAAGAAATGGGCTCCATCTT 697
Qy 541 cctgaacaagtacagaatgaaatcaaatcaatgaagaagaagaatgagtagcattgatt 600
Db 698 CCGGAACAAGTACAGAATGAAATCAAGCAATGAAGAAAGAAATGAGTGAGCTATGATT 757
Qy 601 gattttaaaaaaacctcaatgagatgataccttcccttctgatttcccaaggctgaact 660
Db 758 GACTTCACAAAAACCTCAATGAGGACCATACCTTCCTCTGTATTTCACAGGCTGAATT 817
Qy 661 ggtgctcttcctgatgatttcattgacagtttagaaaaagacagatgatgacaagtataaa 720
Db 818 GGTGCTCTCCTGATGATTTCATTGACAGTTTAGAAAAAGACAGATGACAAAGATATAA 877
Qy 721 attaccttaaaatatcccaactatttccctgtcatgaagaagaatgttgtatccctgaacc 780
Db 878 ATTACTTAAATATCCACACTATTTTCTGTCATGAAGAAATGTTGTATCCCTGAAACC 937
Qy 791 aagaagaagatggaatggcttttaataacaggtgcgaaggaagaaacacataatttg 840
Db 938 AAGAGAAAAATGGAAATGGCTTTTAATACGAGGTGCAAGAGGAAACACCATATCTCTG 997
Qy 841 cagcagctactcccactgcgaaccaaggtgccaaactactcgggttagcacacatgct 900
```

Db 998 CAGGAGCTACTCCACTCGGGGCCAAAGTAGCCAAACTACTTGGGCTACAGCACACATGCT 1057  
Qy 901 gacttcgctcttgaaatgaactgcaaaagacacaaagccgcgtaaacacagcyttcttagat 960  
Db 1058 GACTTTGTCTTGTAGATGAACACTGCGAAGAGATACACACATGTGACAGCCTTCTAGAT 1117  
Qy 961 gatttaagccagaagttaaaacccctgggtgaagcagaacgagagttattttgaatttg 1020  
Db 1118 GATTTAAGCCAGAAATTAACCCCTTGGGTGAGCGCAGACGAGAGTTTCATTTGAATTTG 1177  
Qy 1021 aagaaaaaggaatgcaaaagacaggggttttgaatatgatgggaaaaatcaatgcctgggat 1080  
Db 1178 AAGAAAAAGGAATGTCAGAGAAAGGCTTTTGAATATGATGCGCAAAATCAATGCTGGGAT 1237  
Qy 1081 ctatatctaatgactcagacagaggaactcaaatgatttccatagaccagaaggttccctc 1140  
Db 1238 CTACATTTACTATGACTCAGACAGAAAGCACTCACTGCTGTGGACCAAGAGATCCCTC 1297  
Qy 1141 aaggaatacttcccaattgagtggtcactgaagcgttctgtgaacacacctaccagaggttg 1200  
Db 1298 AAGGAATATTCCTCAATTTGAAGTGTGCTCAATGAGGCTTCTGACACATCTATCAGGAGTTG 1357  
Qy 1201 ttgggacttctattgaaacaaatgacagatgctcatgttttggaaacagaggtttacactt 1260  
Db 1358 TTGGGACTCTCCTTTCAGCAAGTGACAGATGCTCATGTTTGGAAATAAAAGTGTTACCCCTG 1417  
Qy 1261 tatactgtaagagataaagctacagagagaagtattgggacaggttctatttggaccttat 1320  
Db 1418 TACACGGTGAAGGATAAAGCCACAGAGAAAGTATTAGGACAGTTCCTACTTGGACCTCTAT 1477  
Qy 1321 ccaaggaaggaataatcataatcatcgctgctgcttcggtctccagcgtgctgctctctt 1380  
Db 1478 CCNAGGAGAGGAAAAATACAAACATGACGAGCTGCTTGGTCTCCAGCGCTGCTGCCCTCCG 1537  
Qy 1381 cctgatggaagccggatgatggcagtggtgctgctgctgctggtgaaacttctcacagccagtg 1440  
Db 1538 CCTGATGGGAGCGCATGATGCTGTGGTGCCTCCTGCTGTAACCTCTCACAGGCCACGA 1597  
Qy 1441 gcaagtcgctcctctctctgagacacgacgagagtgaggaacttaacttcatgagttggt 1500  
Db 1598 GCAGGTGCGTCCCTCTCTCTGAGACACGATGAGGTGAGGACTTACTTCCATGAATTTGCT 1657  
Qy 1501 cagctgacgacagatttgtgcacagactgattttgcagcagatttagcggaacaaatgtg 1560  
Db 1658 CAGTCTCATGCATCAGATTGTGCGCAGACTGCTTTCGACGATTAGTGAACAAATGTG 1717  
Qy 1561 gaaactgactttgtagagtgccatgcgaactgcttgaataattgggtgtgggacgtcgat 1620  
Db 1718 GAAACTGACTTTGTAGAGTGCCTATCACAAATGCTTTGAAAACTGGGTGTGGGACACTGAT 1777  
Qy 1621 tccctccgaagattgtcaaaacattataaagatggaagccctatttcagacagcatcgtt 1680  
Db 1778 TCCTCCGAAGGCTGTCAAAACATTATAAGATGGAAGGCCCTATTACAGATGATCTGCTT 1837  
Qy 1681 gaaaaacttgttctcttagctgtgtcaacacaggtctcttcgacctgcccagattgtt 1740  
Db 1838 GAAAAACTTGTCTCTAGACTGGTTCACACAGGCTCTCTGACCTTACGCCAGATTGTT 1897  
Qy 1741 ttgagcaaatgtatcagttcttctatacaacacacatcgctggtatgctgcaagtgaatat 1800  
Db 1898 TTGAGCAAAAGTTGATCAGTCTCTCCACACACACACTTCGCTGGATGCTGCAAGTGAATAT 1957  
Qy 1801 gcaaaactgctcagaataattagagttgagcttaactccaggaacaaatgcccagct 1860  
Db 1958 GCCAAATCTGCACAGAAATTTTAGTGTGTGACGTACTCCAGGCCAAATATGCCAGCT 2017  
Qy 1861 acctttggacattttgcaggggatacagatggccaataattatgatattctttggagtgaa 1920  
Db 2018 ACTTTTGGACACTTGGCAGGGGATATGATGGCCATATATTATGATATCTTTGGAGTGAA 2077  
Qy 1921 gtattttccatgatatgttttacagctgttttataaagaagaggaataatgaatccagag 1980  
Db 2078 GTCTTTTCCATGGACATGTTTTATGCTGTTTTTAAAAAAGAGGGATTAATGAATCCAGAG 2137

Qy 1981 gttggaataacatacaaaacctaatctctgaaacctgggggattctctgacgcacatgac 2040  
Db 2138 GTTGGAAATGAATACAGAAACCTTAATCTCTGAAACCTGGGGATCGCTGGACGATGGAC 2197  
Qy 2041 atgctccacaattcttgaacacgtgagcgaacacaaacgcttcctaatgagtagagc 2100  
Db 2198 ATGCTCCAAATTTCTTGAACCGTGAGCCAAACCAAAAGCGTTCTCTAATGAGTAGAGGC 2257  
Qy 2101 ctgcatgctctcgtga 2115  
Db 2258 CTGCATGCTCCGTAA 2272

RESULT 5  
AB000172 2810 bp mRNA MAM - 05-FEB-1999  
LOCUS Porcine mRNA for endopeptidase 24.16, complete cds.  
DEFINITION  
ACCESSION AB000172  
VERSION AB000172.1 GI:1783126  
KEYWORDS endopeptidase 24.16 type M2; endopeptidase 24.16 type M3.  
SOURCE Sus scrofa Adult Liver cDNA to mRNA.  
ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
REFERENCE  
1. (bases 1 to 2810)  
Hirose,S.  
AUTHORS Direct Submission  
TITLE Submitted (27-DEC-1996) to the DDBJ/EMBL/GenBank databases.  
JOURNAL Shigehisa Hirose, Tokyo Institute of Technology, Department of  
Biological Sciences; 4259 Nagatsuta-cho, Midori-ku, Yokohama,  
Kanagawa 226-8501, Japan (E-mail: shirose@bio.titech.ac.jp,  
Tel:045-924-5726, Fax:045-924-5824)  
2. (sites)  
Kato,A., Sugita,N., Saruta,Y., Hosoi,T., Yasue,H. and Hirose,S.  
Targeting of endopeptidase 24.16 to different subcellular  
compartments by alternative promoter usage  
J. Biol. Chem. 272 (24), 15313-15322 (1997)  
97326108 Location/Qualifiers  
1. .2810  
/organism="Sus scrofa"  
/db\_xref="taxon:9823"  
/dev\_stage="Adult"  
/tissue\_type="Liver"  
1. .2810  
/note="corresponding to exon2,5-16 of this gene;  
endopeptidase 24.16 type 2"  
130. .2367  
/standard\_name="endopeptidase 24.16"  
/note="oligopeptidase M :neurolysin :SBAP(soluble  
angiotensin-binding protein) :MEP(microsomal  
metalloendopeptidase)"  
/codon\_start=1  
/product="endopeptidase 24.16 type M2"  
/protein\_id="BAA19063.1"  
/db\_xref="GI:1783127"  
/translation="WVPEGLHRELGLATFSSAPLGGHPFFVWDCLSCKOGDWSQA  
RPTNAERSSGGGIGILLRMTLGRAMSPLOAMSSYTVDRNLKWDLSPOIKRRRT  
ELIAQTQVDDIGMDIEEYENCALADVEKYIVERTMLDPQHVSDEKVR  
AASADKRLSFDIEMSNREDIFLRVLKEDTGVKIKPEARRYLEKSRVGRKRG  
LHLPOVONEIKAMKRMSELCTDFKNLNEEDTFLVFSKALGALPDDFIDSLKTD  
DNKYITLKYPHVFPVWKCCIPETRRKMEMAFNTRKEENTILLOELLPLRAKVAL  
LGYSTHADVLEMTAKSTHHVTAFLLDLSQKLPLGEAREFERILNKKKEBEKFE  
YDGNLWMDLHYMTQTEELKYSDQEIILKEYFPPIEVVTEGLLIYQELLGUSFEQVT  
DAHVKNMSVTLYTKVDKCEVLGVLDLYPREKYNHAACFTGLQPGGLPDGSRMM  
SVAALVWNSQPRAGPSLLRHDEVRTYFHFHVHMQICAOQDFARFSGTNYETDFV  
EYPSOMLENWVNDTSLRLSKHYKDGSPITDLDLEKLVASRLVLTLLTQLQIVLSK  
VQOSLTNTSLDAASEYAKYCTEILGVAATPGTNMPATGCHLAGVDGQYCYLASEV  
FSDMFYSCFKREGIMNPEVKIRNLILKPGSLDGMQLQNLFLAKREPQKAFILMSR  
GLHAP"  
322. .2367  
/standard\_name="endopeptidase 24.16"

mRNA  
CDS

```
/note="alternative product for this mRNA; oligopeptidase M
:neurolysin :SBAP(soluble angiotensin-binding protein)
:MEP(mitochondrial metalloendopeptidase)"
/codon_start=1
/product="endopeptidase 24.16 type M3"
/protein_id="BAA19064.1"
/db_xref="GI:1783128"
/translation="MTLGRKAMSPLOAMSSYTVDRNVLKRWLSPQEIQRRTTEELIAQ
TKQVDDIGMLDIEDIEFVENCLOALADVEVKYIVERTMLDFPQHVSSDKVRAASTEA
DKRLSRFDIEMSRDEIFLRVRLKTCGLDKIKPEARRYLEKSVKMGKRNGLHLPEQ
VQNEKAMKRMSELCDIPNKNLNEDDTFLVSKAEIGALPDDPIDSLEKTDNDKIKI
TLKPYFPVWKCKIPETRRKMEAFNRCKEENTIIQLLELLPURAVALKLGTSYTH
ADFLVEMTAKSTHVTVDLDSOKLPLGEAREFIIQLNKKKECEKEKFEYDCKIN
AWDLHYMTQTEELKYSVDQELKYEFPFIEVTEGLNIIYQELGLSLSEQVTDHVN
KSVLYTVKDKATGEVLDLYPREGKYNHAAFCGLPCLLPDGRSMWSVAALV
VNFSPRAGRSLRLRHDSVRTYEFHEFGVHMQICAOITDFARESGTNTVEFVPSOM
LENWWDPSLRLLKVKDGSPTDDLLLEKLIVASRLVNTGLLRLQIVLSKVQSLH
TNTSLDAASEYAKYKTELLAGVPAITGNPNTFVHLGGYDQGYIYIWSVFSMDMF
YSCFKKEGIMNPEVGMKYRNLILKPGGSLDGMQLNPLKREPNOKFLMSRGLHAP"
BASE COUNT      852 a 558 g 652 g 748 t
ORIGIN

Query Match      86.5%; Score 1829.8; DB 4; Length 2810;
Best Local Similarity 91.7%; Pred. No. 0;
Matches 1933; Conservative 2; Mismatches 174; Indels 0; Gaps 0;

QY 7 gccggtgcttttggctgcgaagctccgcaagttggtggtccagagattttactc 66
DB 259 GCCGCCCCAAAACGAATGACGAAGACGGTCGGGGTGTGGTTCGGGGATTATTG 318
QY 67 agaataagttggaagaagtgatgtctctcttcaggcaatgtcttccatactgtg 126
DB 319 ACAATGACCTTTAGGGAGAGAACCAATGCTCTCTTCAGGCAATGCTTCCTATCTGTG 378
QY 127 gctgcagaaaagtgttttaagatggaattttccaccagagcaaatataaacaagaactgag 186
DB 379 GATGGCAGAAATGTTTAAAGATGGGATCTTTCACCAGAGCAAAATTAAGAAGAACTGAG 438
QY 187 gactcattgtcacacacacaggtacgatgctgttgaatctcgtgattaggaa 246
DB 439 GAGCTGATGGCCGACACCAACAGGTGATGATATATGGGAATCTTGACATTAAGAG 498
QY 247 gtaacttacagaactgtctgcaggcaactgacagatgtagaagtaagtatagtgaa 306
DB 499 GTAACCTATGAGAACTGTTTGACAGCACTGGCAGATGTGGAACTGAATATATAGTGAA 558
QY 307 aggaacatgctagactttccccagcatgtatctctgcacaaagaagtcacgagcagaagt 366
DB 559 CGAAGCATGCTAGACTTTCCTCCAGCATGTCTCCTCTGACAAAGAAGTACGAGCAGCAAGT 618
QY 367 acagacagacaaaagactttctgttttgatattgagatgagatgagagagatatata 426
DB 619 ACAGAAGCAGACAAAAGACTTCTCGTTTGTATATGAGATGAGCATGAGAGAAGATATC 678
QY 427 tttagagaattgttcattacagaaacacctgtgatctggggaagataaaacctgagcc 486
DB 679 TTTCTGAGAAATGTTTCGTTTAAAGAGACCTGTGATTTGGGGAAGATCAAGCCTGAAGCC 738
QY 487 agacgactacttggaagaagtcataatataatggggaagaagaatgggctccattcttgaa 546
DB 739 AGAAGATACTTAGAAAAATCAAGTTTAAATGGGAAAAAAGAAATGGGCTCCATCTCCGGAA 798
QY 547 caagtacagaatgaatcaaatcaatgaagaaaagaatgaagtgaactgattgatttt 606
DB 799 CAAGTACAGAATGAATCAAGCAATGAGAAAGAAATGAGTGAGCTATGATTAAGACTTC 858
QY 607 acaaaaaacctcaatgagatgataccttctgttatttccaaagctggaactgtgtgct 666
DB 859 AACAAAAACCTCAATGAGGACGATACCTTCCTGTGATTTTCCAAAGCTGAACCTGTGCT 918
QY 667 cttctcgtgattttcattgacagtttagaagaacagatgatgacaagtataaaattacc 726
```

```
DB 919 CTTCTCTGATGATTTCATTGACAGTTTAGAAAAAGACAGATGACAAAGTATATAAATTACC 978
QY 727 ttaaatatccacacatttccctgtcatgaagaatgttgtatccctgaaaccagaaga 786
DB 979 TTAANAATATCCACATATTTTCTGTGATGAAGAAATGTTGTATCCCTGAACACAGAAG 1038
QY 787 aggatggaatggctttttaatacaaggtgcaagaggaagaaacacacataattttgcagag 846
DB 1039 AAAATGGAAATGGCTTTTAAATACGAGGTGCAAAAGAGAAACACCATATATCTCGCAGGAG 1098
QY 847 ctactcccaactgcgaaccaagtgccaaaacttactcgtttatagcacacactcgtgacttc 906
DB 1099 CTACTCCCACTCGGGGCCAAAGTACGCAAACTACTTGGGTACGACACACATGCTGACTTT 1158
QY 907 gctcctgaaatgaacactgcgaagagcacaagccgtaacagcgttttctagatattta 966
DB 1159 GTCCCTTGAGATGAACACTGCGGAAGAGTACACACCATGTGACAGCCCTTCTAGATGATTTA 1218
QY 967 agccagaagttaaaaaccccttgggtgaaagcagacgagagttattttgaatttgaagaaa 1026
DB 1219 AGCCAGAAATTAACCCCTTGGGTGAGGCAAGACGAGAGTTTCATTTTGAATTTGAAGAAA 1278
QY 1027 aaggaatgcaaaagacaggggttttgaatatatgatgggaaaaatacaatgcctgggatatata 1086
DB 1279 AAGGAATGTGAAGAAAGAGTGTTCGAATATGATGCAAAATCAATGCCCTGGGATCTACAT 1338
QY 1087 tactacatgactcagacagagaaactcaagtattccatagaccagagttctcctaagaa 1146
DB 1339 TACTATATGACTCAGACAGAGAACTCAAGTACTCTGTGGACCAAGAGATCTCTCAAGGNA 1398
QY 1147 tacttcccaattgagtggtcactgaaaggcttgcctgaacacactaccagaggttgttgga 1206
DB 1399 TACTTCCCAATTAAGTGGTCACTCAAGGCTTGGCTGAACATCTATCAGGAGTGTGTGGGA 1458
QY 1207 ctttcatttgaacaaatgacagatgctcatgttttgaacaaagttgtacactttatact 1266
DB 1459 CTCCTCTTGGAGAGTGCACATGCTCATGTTTGGAAATTAAGTGTTCACCTGTACACG 1518
QY 1267 gtgaagataaagctcacaggaagatttggacagttctatttggacctctatccaagg 1326
DB 1519 GTGAAGGATAAAGCCACAGGAGAAATATTAGACAGATTCTTACTTGGACCTCTATCCAAG 1578
QY 1327 gaagaaaaatacaatctgcgcctgcttgcgtctccagcctggctgcttctgctgat 1386
DB 1579 GAAGGAANAATACAACCATGCACCTGCTTGGTCTCCAGCCTGGCTGCTCTGCTGCTGAT 1638
QY 1387 ggaagccgaatgtaggcagtggtgctccctctggttgaacttctcacagccagtgaggt 1446
DB 1639 GGGAGCCGATGATGTCGTGTGGCTGCCCTCGTGTGGTGAACCTTCTCACAGCCACGAGCAGT 1698
QY 1447 cgtccctctcttgagacagacagaggtgaggaacttactttcagtttgggtcacatg 1506
DB 1699 GCTCCTCTCTCCTGAGACAGATGAGGTGAGGACTTACTTCCATGAATTTGGTCACTGTC 1758
QY 1507 atgcatcagatttgcacagactgattttgcacgatttagcggaacaaaatgtggaact 1566
DB 1759 ATGATCAGATTTGTGGCCAGACTGACATTTGCACGATTTAGTGGAAACAAATGTGGAACAT 1818
QY 1567 gactttagaggggtccatcgcaaatgcttgaataatgggtgtgggacgtcgattccctc 1626
DB 1819 GACTTTGTGAGAGTGGCCATCAAAATGCTTGAATACTGGGTGGGACACTGATTTCCCTC 1878
QY 1627 cgaagatgtcaaaacattataaagatggaagccctatttgcagacgactctgcttgaagaaa 1686
DB 1879 CGAAGCTGTCAAACATTTAAAGATGGAAGCCCTATTACAGATGATCTGCTTGAAGAAA 1938
QY 1687 cttgtgctttaggtgttcaacacaggttcttgcacctgctgacctgacctgatttttggagc 1746
DB 1939 CTGTGTGCTTCTAGACTGGTCAACACAGGCTCTCCTGACCTTACGCCAGATGTTTGTGAGC 1998
QY 1747 aaagttagatgctcttccatcaccaacacatcgctgagatgctggaagtgaatataagccaaa 1806
DB 1999 AAAGTTGATGCTCTCTCCACACCAACACTTTCGCTGGATGTGCAAGTGAATATATGCCAAA 2058
```

QY	1807	tactgctcagaataatattagagttgcagctactccaggccacaaatgacagctacattt	1866			
Db	2059	TACTGCACAGAAATTTAGTGTTCAGCTACTCCAGGCACAAATATGCCAGCTACTTTT	2118			
QY	1867	ggacatttggcggggatgacatgagccaattattgatatcttcttgagtgaaagtattt	1926			
Db	2119	GGACACTTGGCAGGGGATATGATGGCCAAATATTATGATATCTTTGGAGTGAAGTCTTT	2178			
QY	1927	tcacatgatattttcacagctgttttaaaaaagaagggaataatccagaggttga	1986			
Db	2179	TCCATGGACATGTTTATAGCTGTTTAAAGAAAGAGGGAATAATGAATCCAGAGGTGGA	2238			
QY	1987	atgaatacagaaacctaactcctgaaacctggggatctctgacggcatgacatgctc	2046			
Db	2239	ATGAATATACGAACCTTAATCTCTGAACCTGGGGATCCTGGACGGCATGACATGCTC	2298			
QY	2047	cacaattttgaaacgtgagccaaaccaaagcgttctcctaagttagagggcctgcatt	2106			
Db	2299	CAAAATTTCTTGAACGTGAGCCAAACCAAAAGCCTTCTTAATGATGATAGAGGCTGCAT	2358			
QY	2107	gctyctgta	2115			
Db	2359	GCTCCGTAA	2367			
RESULT	6					
AB000171						
LOCUS	AB000171	2709 bp	mrna			
DEFINITION	Porcine mRNA for endopeptidase 24.16, complete cds.					
ACCESSION	AB000171					
VERSION	AB000171.1	GI:1783124				
KEYWORDS	endopeptidase 24.16 type M3.					
SOURCE	Sus scrofa Adult Liver cDNA to mrna.					
ORGANISM	Sus scrofa					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.					
AUTHORS	Hirose, S.					
TITLE	Direct Submission					
JOURNAL	Submitted (27-DEC-1996) to the DDBJ/EMBL/GenBank databases.					
	Shigeisha Hirose, Tokyo Institute of Technology, Department of Biological Sciences; 4259 Nagatsuta-cho, Midori-ku, Yokohama, Kanagawa 226-8501, Japan (E-mail:shiroyse@bio.titech.ac.jp, Tel:045-924-5726, Fax:045-924-5824)					
REFERENCE	2 (sites)					
AUTHORS	Kato, A., Sugiura, N., Saruta, Y., Hosoi, T., Yasue, H. and Hirose, S.					
TITLE	Targeting of endopeptidase 24.16 to different subcellular compartments by alternative promoter usage					
JOURNAL	J. Biol. Chem. 272 (24), 15313-15322 (1997)					
MEDLINE	97326108					
FEATURES	Location/Qualifiers					
source	1..2709					
	/organism="Sus scrofa"					
	/db_xref="taxon:9823"					
	/dev_stage="Adult"					
	/tissue_type="Liver"					
mrna	1..2709					
	/note="corresponding to exon 1,4,5-16 of this gene; endopeptidase 24.16 type 1"					
CDS	221..2266					
	/standard_name="endopeptidase 24.16"					
	/note="oligopeptidase M: neurolysin :sBAP(soluble angiotensin-binding protein) :MEP(mitochondrial metalloendopeptidase)"					
	/codon_start=1					
	/product="endopeptidase 24.16 type M3"					
	/protein_id="BAA19062.1"					
	/db_xref="GI:1783125"					
	/translation="MTLGRAMSPLQAMSSVTVDGRNVLRWDLSPQIKRTEELIAQ TKQVDDIGMLDIEVTYENCLOALADVEKYIVERTMLDFPQHVSSDKVRAASTEA DKRLSRFDIEMSRDIEDLRIVRLKTCDLGKIKPEARRYLEKSVKMGKRNGLHLPEQ VQNEIKAMKRMSELICDFKNLNEBDDTFLVFSKALGALPDFIDLSLEKTDNDNKYKI					
BASE COUNT	826 a	532 c	616 g			
ORIGIN	735 t					
Query Match						
Best Local Similarity 86.4%; Score 1828.4; DB 4; Length 2709;						
Matches 1922; Conservative 2; Mismatches 158; Indels 0; Gaps 0;						
QY	34	ctccgcagagttggtgtccaggattttactcagaatgacgttaggaagagaagtgatg	93			
Db	185	CACATCTGGGTTGGTTCGGGATTTTATTCAGAAATGACTTTTAGGAGAGAAGCAATG	244			
QY	94	tctcctctccaggaatgctctcctatactgtggctggcagaaatgttttaagatggat	153			
Db	245	TCCTCTCTTCAGGCAATGCTTCCTATCTATCTGTGTGGATGAGAAATGTTTAAAGATGGGAT	304			
QY	154	ctttccaccagagcaaatataacaaagaaactgagagctcattgtgcagaccacaacaggtg	213			
Db	305	CTTTCACAGAGACAAATTAAGAAGAACTGAGGAGCTGATTCGCGCAGACCAACAGGTG	364			
QY	214	taogatgtgttggaatgctcgtgatttgagggaagttaactacagaaactgtctgcaggaa	273			
Db	365	TATGATGATATTGGAATGCTTGACATTTGAAGAGGTAACCTTATGAGAACTGTTTGCAGGCA	424			
QY	274	ctggcagatgtagaactaaagtatatagtggaagagacatgctacttccccagcat	333			
Db	425	CTGGCAGATGTGGAAGTGAAGTATATAGTGAACAGCAACTGCTAGACTTTCCTCCAGCAT	484			
QY	334	gtatcctctgcacaaagaactacgagcagcaagtacagaagcagacaaagacttctcgt	393			
Db	485	GTCTCTCTGACAAAGATACGAGCAGCAGCAAGTACAGAGCAGACAAAGACTTTCCTCGT	544			
QY	394	tttgattgagatgagcagcagagagagatatatttgagagaattgttcatttcacaggaa	453			
Db	545	TTTGATATTGAGATGAGCATGAGAGAAGATATCTTCTGAGAAATGTTCTGTTTAAAGGAG	604			
QY	454	acctgtgatctgggaagataaaacctgaggcagcagcacttggaagaaagtcacaaataa	513			
Db	605	ACCTGTGATTTTGGGAAGATCAAGCCTGAAGCCAGAGATGATTTAGAAAATCAGTTAAA	664			
QY	514	atggggaagaaatgggctccatctctgaacaaagtcagaaatgaaatcaaatcaatg	573			
Db	665	ATGGGAAAAGGATGGGCTCCATCTTCGGAACAGTACAGAAATGAAATCAAGCAATG	724			
QY	574	aagaaagaatgagtgagctatgtattttaaacaacaaacctcaatgaggatgatacc	633			
Db	725	AAGAAAAGAAATGAGTGAGCTATGATTGACTTCAACAAAAACCTCAATGAGGACGATACC	784			
QY	634	tcccttgatttttccaaagctgaactggtgctctctccatgatttccattgcagattta	693			
Db	785	TTCTCTGTATTTTCCAAAGGCTGAACCTGGTGTCTTCTCTGATGATTTCATGTGACAGTTTA	844			
QY	694	gaaagacagatgatgacaagtataaaataaccttaaaataatccacactatttccctgtc	753			
Db	845	GAAAACAGATGACACCAAGTATAAATACCTTAATATATCCACACTATTTCCTGTC	904			
QY	754	atgaagaatgttgtatccctgaaacacagagaagtgaaatggcttttaatacaagg	813			
Db	905	ATGAGAAAATGTTGTATCTCCTGAAACACAGAGAAAATGAAAATGGCTTTTAAATACGAGG	964			
QY	814	tgcagaagagaaacacataattttgacagcactctcccactggaacaaaggtgccc	873			
Db	965	TGCAAGAGAAAACACCAATAATCTCCAGGAGCTACTCCACTGCGGGCCCAAGTAGCC	1024			
QY	874	aaactactcggttatagcacacatgctgacttcgttccttgaaatgaacatgcaagagac	933			

Db 1025 AAACCTACTGGGCTACAGCACACATGCTGACTTTGTCTTGTGAGATGAACACATGCGAAGAGT 1084  
Qy 934 acaagccgcgttaacagcytttctagatgatttaagcagaagagtttaaaacccottgggtgaa 993  
Db 1085 ACACACCATGTCACAGCCCTTCTAGATGATTAAAGCCAGAAATTAACACCCCTGGGTGAG 1144  
Qy 994 gcagacagagttatttttaatttgaagaaaggaatcaagacagaggggttttgaa 1053  
Db 1145 GCAGAACGAGAGTTCATTTTGAATTTGAAGAAAGAGGATGTGAAGAAAGGTTTTGAA 1204  
Qy 1054 tatgtggaaacaatcgaatgcctgggatctatttactactgactcagacagagaaactc 1113  
Db 1205 TATGATGCAAAATCAATGCGTGGATCTACATTAATATGACTCAGACAGAAAGAACTC 1264  
Qy 1114 aagtattccatagaccaagagttcctcaaggaatacttcccaattgaggtggtcactgaa 1173  
Db 1265 AAGTACTCTGTGGACCAAGAGATCCTCAAGGAATACTTTCCCAATTTGAAGTGTGCTACTGAA 1324  
Qy 1174 ggtctgtgaacactaccagaggtgttgagactttctatttgaacaaatgacagatgct 1233  
Db 1325 GCCTGTGCAACATCTATCAGAGTGTGTGGACTCTCCTTTGACCAAGTGCACAGATGCT 1384  
Qy 1234 catgtttggaacaagagttgtacactttatactgtgaaggataaagctcacaggagaagta 1293  
Db 1385 CATGTTTGAATAAAGGTAACTACCTGTACACGGTGAAGGATAAAGCCACAGGAGAAGTA 1444  
Qy 1294 ttgggacagttctatttggacctctatcccaagggaagaaatacaatcatctgcggcctgc 1353  
Db 1445 TTAGGACAGTTCTACTTTGGACTCTATCCAAAGGAAGGAAATACACATTTGACGACTGC 1504  
Qy 1354 tctggtcccaagcctgcctctctgctgctgtaggaacggagtagcagtgctgctgc 1413  
Db 1505 TTTGGTCTCCAGCCTGGCTGCTGCTGCTGATGGGAGCCGACGATGCTGTGTGGCTGCC 1564  
Qy 1414 ctctggtggaactctcaacagccagtggtgagacttctctctcctgagacacagcag 1473  
Db 1565 CTCGTGTGTGAATCTCTACAGCCACGACGAGCTGCTCCTCTCTCTGAGACAGATGAG 1624  
Qy 1474 gtgaggacttacttcatgagtttggtaacgtgatgatacagatttggacagactgat 1533  
Db 1625 GTGAGGACTTACTTCCATGAATTTGGTACACCTCATGCTATGATGTTGTGCGACAGCTGAC 1684  
Qy 1534 ttgcacgatttagcgaacaaatgtgaaactgacttctgtagagtgccatgcacaaatg 1593  
Db 1685 TTTGCACGATTTAGTGGACAAATGTGGAATGACTTTGTGTAGAGTGCCATCACAATG 1744  
Qy 1594 cttgaaattgggtgtggagcgtgattccctccgaagattgtcaaaacattataaagat 1653  
Db 1745 CTTGAAAACCTGGGTGTGGACACTGATTCCTCCGGAAGGCTGTCAAAACATATATAAGAT 1804  
Qy 1654 ggaagccctattgcagacgactctgtgaaacttggcttctgcttagctggtcagacaca 1713  
Db 1805 GGAAGCCCTATTACAGATGATCTGCTTGAAAACTTGTGTCTCTAGACTGGTCAACACA 1864  
Qy 1714 ggtctctgacccctgcgcagattgttttgagcaaatgtatcgactctctcattaccaac 1773  
Db 1865 GGTCTCTGACCTTACGCCAGATGTTTGTGAGCAAGTTGATGATGATCTCTCCACACCAAC 1924  
Qy 1774 acatcgctgagtgctgcaagtgaatatgccaaaactactgctcgcagaataattaggagttgca 1833  
Db 1925 ACTTCGCTGGATGTGCAAGTGAATATGCCAAATACTGTCACAGAAATTTTAGGTGTGCA 1984  
Qy 1834 gctactccagacaaatatgcagctacccttggacaatttggacatttggcaggggatacagatggc 1893  
Db 1985 GCTACTCCAGGCAAAATATGCCAGCTACTTTTGGACATTTTGGCAGGGGGATATGATGGC 2044  
Qy 1894 caattattgatatcttggagtgaaagtatttccatggatatttttccagctgtttt 1953  
Db 2045 CAATATTATGATATCTTTGGAGTGAAGTCTTTTCCATGGACATGTTTATTAGCTGTTT 2104  
Qy 1954 aaaaaaaggggataatgaatccagaggttggatgaatacagaaacctaatactctgaa 2013

Db 2105 AAAAAAGAGGATAATGAATCCAGAGTTGGAATGAATAACAGAAACCTAATCCTGAAA 2164  
Qy 2014 cctgggagatctctggacgcatgacatgctccacatttcttgaacgtgagccaaac 2073  
Db 2165 CCTGGGATCGCTGGAGCGCATGACATGCTCCAAATTTCTTGAACAGTGAGCCAAAC 2224  
Qy 2074 caaaaagcgttccaaatgagtagagcctgcatgctcgtcgtga 2115  
Db 2225 CAAAAGGTTCTTAATGAGTAGAGGCTGTCATGCTCCGTAA 2266  
RESULT 7  
AB000173 2862 bp mRNA MAM 05-FEB-1999  
LOCUS Porcine mRNA for endopeptidase 24.16, complete cds.  
ACCESSION AB000173  
VERSION AB000173.1 GI:1783129  
KEYWORDS endopeptidase 24.16 type M3.  
SOURCE Sus scrofa Adult tissue\_lib:Liver cdna to mRNA.  
ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
REFERENCE 1 (bases 1 to 2862)  
AUTHORS Hirose, S.  
TITLE Direct Submission  
JOURNAL Submitted (27-DEC-1996) to the DDBJ/EMBL/GenBank databases.  
Shigehisa Hirose, Tokyo Institute of Technology, Department of  
Biological Sciences; 4259 Nagatsuta-cho, Midori-ku, Yokohama,  
Kanagawa 226-8501, Japan (E-mail: shirose@bio.titech.ac.jp,  
Tel: 045-924-5726, Fax: 045-5824)  
REFERENCE 2 (sites)  
AUTHORS Kato, A., Sugiyama, N., Saruta, Y., Hosoiri, T., Yasue, H. and Hirose, S.  
TITLE Targeting of endopeptidase 24.16 to different subcellular  
compartments by alternative promoter usage  
J. Biol. Chem. 272 (24), 15313-15322 (1997)  
JOURNAL MEDLINE  
FEATURES 97326108  
Location/Qualifiers  
source 1. .2862  
/organism="Sus scrofa"  
/db\_xref="taxon:9823"  
/dev\_stage="Adult"  
/tissue\_lib="Liver"  
1. .2862  
/note="corresponding to exon2.4.5-16 of this gene;  
endopeptidase 24.16 type 2"  
374. .2419  
/standard\_name="endopeptidase 24.16"  
/note="oligopeptidase M :neurolysin :sBAP(soluble  
angiotensin-binding protein) :MEP(microsomal  
metalloendopeptidase)"  
/codon\_start=1  
/product="endopeptidase 24.16 type M3"  
/protein\_id="BAAI9065.1"  
/db\_xref="GI:1783130"  
/translation="MTLGREAMSPLOAMSSYTVDGRNVLRLWDLSPQIKRRTBELIAQ  
TKGVYDDIGMLDIEEYTVGRCALADVEVKYIVERTMLDFPQHVSSDKKVRRASTEA  
DKSLRFDIEMSRMEDIFLRLVKETCDLGKIKPEARVLEKSKVRKMGKRLHPQ  
VONEIKAMKRMSELCTIDFNKLNEDDTFLVSKAELGALPDDFIDSLKTDNDKVKI  
TLKYPHPVKKCCIPETRRKMEMAFNCKEENTIIQLLEPLRAKVAKLLGYSTH  
ADPVEMLNTAKSTHHVTAFLDDUSQKLPLGEAREFILLNKKKECEKEGFEIDGKIN  
ANDHYMTQTTELKYSDQETLKEYFPIEVVTEGLINIQELLGLSFEQVTDHVVN  
KSVTLTYTKDKATGEVLGYLDLYPREGYNHAACGFLQPLGCLLPDGSRRMSVAALY  
VNEFSQPRAGPSLLRHDEVYTFEFGVHMQICAOQDFARFGTNNVETDFEVSOM  
LENWVNDTSLRLSKHYKDGSPITDQLLEKLVASRLVNTGLLTLRQIVLSKVDOSLH  
TNTSLDAASEYAKYCTEILGVAATPGTNMPATFCHLAGGYDGYGYGLWSEVFSMDHF  
YSCFKREGINPNPVGMRKRLNLIKPGSGSDGMDMLQNLKREPKNQAFILMSRGLHAP"  
BASE COUNT 866 a 568 c 656 g 772 t  
ORIGIN  
Query Match 86.4%; Score 1828.4; DB 4; Length 2862;  
Best Local Similarity 92.3%; Pred. No. 0;  
Matches 1922; Conservative 2; Mismatches 158; Indels 0; Gaps 0;

QY 34 ctccgcagagtggtggtccaggtatttactcacaatgacgttaggaagaaagtatg 93  
DB 338 CACATCTGGTGGTGGTCCGGGATTTTATGAGAAAGACTTTAGGGAGAGAAGCAATG 397  
QY 94 tctcctcttcaggcaatgctcttcctactactggtggtgcagaaaatgttttaagatggat 153  
DB 398 TCTCCTCTTCAGGCAATGCTTCTCTATCTGTTGGATGGCAGAAATGTTTTAAGATGGGAT 457  
QY 154 ctttcaccagagaacaattaaaacagaactgaggagctcaattgtgcagacccaacagggtg 213  
DB 458 CTCTCACCCAGAGCAAAATTTAAAGAGAACTGAGGAGCTGATTTGCGCAGACCAACAGGTG 517  
QY 214 tacgatgctgttggagtgctgcgtttaggaagtaacttacgagaactgtctgcaggca 273  
DB 518 TATGATGATATTGGGAATGCTTACATTTAGAGAGGTAACTTATGAGAACTGTTTGAGGCA 577  
QY 274 ctggcagatgtagaagtaagatatagtggaagagaccatgtagaactttcccagcat 333  
DB 578 CTGGCAGATGCGAAAGTGAAGTATATAGTGAACGAAACCATGCTAGACTTTCCCCAGCAT 637  
QY 334 gtatcctctgacaagaagtacgagcagcaagtacagaagcagacaaaagaactttctcat 393  
DB 638 GTCTCCTCTGACAAAGAGTACGACACCAAGTACAGAGCAGACAAAGAAAGACTTTCTCT 697  
QY 394 ttgtgattgagatgagcatgagagagatatatttgagagaattgttcatattacaggaa 453  
DB 698 TTTGATATTGAGATGAGCATGAGAGAGATATCTTCTGAGAAATGTTCTGTTAAAGGAG 757  
QY 454 acctgtgatctgggaagataaaaacctgagccagacacatacttggaaaagtcaattaaa 513  
DB 758 ACCTGTGATTTGGGGAAGTCAAGCTGAAGCCAGAAGTACTTAGAAAAATCAGTTAAA 817  
QY 514 atggggaaaagaatgggtccatctctcctgaaacaggtacagaatgaaatcaaatcaatg 573  
DB 818 ATGGAAAAAGAAATGGGCTCCATCTTCGGAAACAAGTACAGAATGAATCAAAACAATG 877  
QY 574 aagaaaagaatgagtgagctatgtattgattttaacaaaacctcaatgagatgatacc 633  
DB 878 AGAAAAAGAAATGAGTGAGTATGTATGACTTACAAAAAACCCTCAATGAGGACGATACC 937  
QY 634 ttccctgtatttccaaggctgaacttggtgctctctctgtatgatttcattgacagtta 693  
DB 938 TTCTCTGTATTTTCCAAGGCTGAACCTTGCTGCTCTTCTCTGATGATTTTCATTACAGTTTA 997  
QY 694 gaaagacagatgtagaagataaaaataccttaataataatccacactatttccotgtc 753  
DB 998 GAAAAACACAGATGACAACAAGTATAAAATTACCTTAAAAATATCCACACTATTTTCTCTGTC 1057  
QY 754 atgaagaatgtgtatccctgaaaccagaagaagatggaatggcttttaatacaagg 813  
DB 1058 ATGAAGAAATGTTGTATCCCTGAAACCAGAGAAATGGAATGGCTTTTAAATACGAGG 1117  
QY 814 tgcaagaggaataacacataaatttgacagcgtactcccactcgcaaccaagtggtcc 873  
DB 1118 TCCAAAGAGAAAAACCATATATCTCGAGGAGCTACTCCCACTGGGGCCCAAGTAGCC 1177  
QY 874 aaactactggtttatagcacacatgctgactctgctctcttggaaatgaacactgcaaaagac 933  
DB 1178 AAATCTACCTGGCTACAGCACACATGCTGACTTTGCTTTGAGATGAACATCGGAAGAT 1237  
QY 934 acaagccgcgtacaagcyttcttagatgatttaagccagaagttaaaaccttgggtgaa 993  
DB 1238 ACACACCATTTGACAGCCCTTCTAGATGATTTTAAAGCAGAAATTTAAACCCCTTGGGTGAG 1297  
QY 994 gcagacagagatttatttgaattgaagaaaggaatgcaagacaggggttttga 1053  
DB 1298 GCAGAACGAGAGTTCATTTTGAATTTGAAGAAAAAGGAATGTGAAGAGAAAGGTTTGA 1357  
QY 1054 tatgatgggaaaaatcaatgcctgggtatctatatattactactgactcagacagaggaactc 1113  
DB 1358 TATGATGGCAAAATCAATGCCCTGGGATCTACATTACTATATCTACTCAGACAGAAACATC 1417

QY 1114 aagttatccatagacacaagagtgctctcaaggaaatacttcccattgaggtggtcaactgaa 1173  
DB 1418 AAGTACTCTGTGGACCAAGAGATCTCAAGGAATACTTCCCAAATTTGAAGTGTCACTGAA 1477  
QY 1174 ggtctgtgaacacacatccaggaggtgttggtggaacttcaatttgaacaaatgacagatgc 1233  
DB 1478 GGCTTGTCTGAACATCTATCAGGAGTGTGTGGGACTCTCTCTTTGAGCAAGTGACAGATGCT 1537  
QY 1234 catgtttggaaacaagagtggttacactttatactgtgaagataaaagctacagagagaaga 1293  
DB 1538 CATGTTTGGAAATAAAAGTGTATACCGGTACACGGTGAAGGATAAAAGCCACAGAGAAGTA 1597  
QY 1294 ttggacaggttctatttggacctctatccaagggaaggaatacaaatcatcgccctgc 1353  
DB 1598 TTAGACAGATTCTACTTGGACCTCTATCCAGGAGGAGAAATACACCATCGAGCTGC 1657  
QY 1354 ttcggtctccagcctggtgcttgcctgtatggaagccggatgatggtcagtggtgcgc 1413  
DB 1658 TTTGGTCTCCAGCCTGGCTGCTCTGATGGAGCGCATGATGCTCTGTGGCTGCC 1717  
QY 1414 ctctgtgtgaacttctcacagccagtggaagctgtccctctctctctgagacagcagag 1473  
DB 1718 CTGCTGGTGAACCTTCTACAGCCACGAGCAGGTGCTCTCTCTCTGAGACCGATGAG 1777  
QY 1474 gtgagacttactttcatgagtttggtaacgtgacatcagatattgtgcacagactgat 1533  
DB 1778 CTGAGGACTTACTTCCATGAATTTGGTCACTGTCATGCATCAGATTGTGCGGACAGTGA 1837  
QY 1534 ttgtcaacgatttagcgggaacaaatgtggaactgactttgttagaggtgacctgcgaatg 1593  
DB 1838 TTTGACAGATTTAGTGGAAACAAATGTGAAACTGACTTTGTAGAGGTGCCATCACAAATG 1897  
QY 1594 cttgaaaattgggtgtggagctgacttccctccgaagattgtcaaaacattataaagat 1653  
DB 1898 CTGAAAACTGGGTGTGGGACACTGATTCCTCCGAAGCTGTCAAAACATTATAAAGAT 1957  
QY 1654 ggaagccctattgcagaagatctgtgaaaactgttgcctcttaggtgcacacaca 1713  
DB 1958 GGAAGCCCTATTACAGATGATCTGCTTGAANAACCTTGTGCTTCTAGACTGCTCAACACA 2017  
QY 1714 ggtctctgacacctgcccagatgtttttgagcaagttgatacagtcctctctcattacaac 1773  
DB 2018 GGTCTCTGACCTTACGCCAGATTTGTTTGAAGTGTGATCAGTCTCTCCACACCAAC 2077  
QY 1774 acatcgctgagatgctgcaagtaataatgcaaaataactgctcagaataattagagattga 1833  
DB 2078 ACTTCGTGGATGCTGCAAGTGAATATGCAAAATCTGCACAGAAATTTTAGGTGTGCA 2137  
QY 1834 gctactccaggcacaaaatatgccacttgcacatttgcaggggggatacagatgc 1893  
DB 2138 GCTACTCCAGGCACAAATATGCCAGTACTTTTGGACACTTGGCAGGGGATATGATGCG 2197  
QY 1894 caatatattgatatctttggagtgaaagtattttccatggatgtttttcacagctgtttt 1953  
DB 2198 CAATATTATGGATATCTTTGGAGTGAAGTCTTTTCCATCGACATGTTTATAGTGTCTTT 2257  
QY 1954 aaaaagaaggggataatgaatccagaggttggaatgaatacagaaaacctaactcctgaaa 2013  
DB 2258 AAAAAAGAGGGATAATGAATCCAGAGGTGGAATGAAATACAGAAACCTTAATCTCTGAAA 2317  
QY 2014 cctgggggatactctggagcgaatgacatgctcccaaatcttcttgaacgtgagcaaac 2073  
DB 2318 CTCTGGGGATCTCTGGAGCGCATGACATGCTCCAAAATTTCTTGAACAGTGAGGCCAAC 2377  
QY 2074 caaaaagcgttcttaagttagtaggcctgcactgctgcgtga 2115  
DB 2378 CAAAAGCGTTCCTTAATGAGTAGGCCCTGCATGCTCCGTAA 2419

RESULT 8

AB000174

LOCUS AB000174 2604 bp mRNA

MAM

05-FEB-1999

DEFINITION Porcine mRNA for endopeptidase 24.16, complete cds.



AB000174  
VERSION AB000174.1 GI:1783131  
KEYWORDS endopeptidase 24.16 type M3; endopeptidase 24.16 (Oligopeptidase M)  
SOURCE Sus scrofa Adult Liver cDNA to mRNA.  
ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
REFERENCE 1 (bases 1 to 2604)  
AUTHORS Hirose, S.  
TITLE Direct Submission  
JOURNAL Submitted (27-DEC-1996) to the DDBJ/EMBL/GenBank databases.  
Shigehisa Hirose, Tokyo Institute of Technology, Department of  
Biological Sciences, 4259 Nagatsuta-cho, Midori-ku, Yokohama,  
Kanagawa 226-8501, Japan (E-mail: shirose@bio.titech.ac.jp,  
tel:045-924-5726, fax:045-924-5824)  
REFERENCE 2 (sites)  
AUTHORS Kato, A., Sugiyama, N., Saruta, Y., Hosoliri, T., Yasue, H. and Hirose, S.  
TITLE Targeting of endopeptidase 24.16 to different subcellular  
compartments by alternative promoter usage  
J. Biol. Chem. 272 (24), 15313-15322 (1997)  
MEDLINE 97326108  
FEATURES  
source 1..2604  
    /organism="Sus scrofa"  
    /db\_xref="taxon:9823"  
    /dev\_stage="Adult"  
    /tissue\_type="Liver"  
    1..2604  
    /note="corresponding to exon3,5-16 of this gene;  
    endopeptidase 24.16 type 3"  
    116..2161  
    /standard\_name="endopeptidase 24.16"  
    /note="oligopeptidase M: neurolysin :sBAP(soluble  
    angiotensin-binding protein) :MEP(mitochondrial  
    metalloendopeptidase)"  
    /codon\_start=1  
    /product="endopeptidase 24.16 type M3"  
    /protein\_id="BAAL9066.1"  
    /db\_xref="GI:1783131"  
    /translation="MTLGRAMSPLOAMSSYTVDRNVNLDWLSPEQIKRTEELIAQ  
    TKQYDIDGMDIEVTENCLQALADVEVYIVERTNLDFFQHVSSDKVRAASTBA  
    DKRLSDIFEMSRNEDIFLRIVRLKTCDLGKPKPARYLEKSVKMRGNLHLPQ  
    VONEIKAMKRMSCLDFNKNLNEDDTFLVSKAELGALPDFFDSLEKTDNDKIKI  
    TLKYPFVPMKKCIPTFRKMEMAFNTRCKEENTIIQLLEPLRAKVAKLIGSYH  
    ADFLVNTAKTHVTAFLDLSOKLPLGEAREFILLKKKECEKGFVDGKTN  
    AWDLHYVTQTEELKYSVDQELIKYEPFIEVTVTEGLNIYQELGLSFEQVTDHVN  
    KSVTLIYVKAKAGEVLQYLDLYPREKYNHACRGLQPGCLLPDGRMSVAALY  
    VNFQPRAGPSLLRHDEVRTYFHEGHVMHQCAQDFARFSGTNVETDFVEVPSOM  
    LENWVPTDLSRLSKHYKDGSPITDLDLEKLVASRLVNTGLLTLRQIVLSKVDQSLH  
    TNTSLDASEYAKYCTEILGVAATPGTNMPATFGLAGYDGYGYGLWSEVFSMDMF  
    YSCFKKEGIMNPEVGMKRYRNLIKPGSLDMDMLQNLKRPKNQKAFKMSRGLHAP"  
BASE COUNT 820 a 478 c 594 g 712 t  
ORIGIN  
  
Query Match 86.4%; Score 1828.2; DB 4; Length 2604;  
Best Local Similarity 92.4%; Pred. No. 0;  
Matches 1920; Conservative 2; Mismatches 155; Indels 0; Gaps 0;  
  
Qy 39 cagagtggtggtccagattttactcagatgacgttagaagagagagtgatgtctcc 98  
Db 85 CGGGTGTGGTTCGGGATTTTATTGAGAATGACTTTAGGAGGAAGCAATGTCTCC 144  
Qy 99 tcttcaggaatgcttctactgtgtgctggcagaaatgttttaagatggatcttcc 158  
Db 145 TCTTCAGGAATGCTTCTCTACTGTGGATGGCGAGAAATGTTTAAGATGGGATCTTC 204  
Qy 159 accagagcaaatataaacaagaactgagagagctcattgtgcagacaaacaggtgtacga 218  
Db 205 ACCAGAGCAAAATTAAGAGAACTAGGAGCTGATTCGGCAGACCAACAGGTGTATGA 264  
Qy 219 tgcgttggaatgctcgtgtattgaggaagtaacttaacgaagactgtctgcaggcactggc 278



Db 1345 ACAGTTCTACTGGACCTCTATCAACAGGGAAGAAATAACAACATCAACGCCCTGCTTTGG 1404

Qy 1359 tctccagctggctgctctctgctgctgagcagcgatgagcagtgctgctcctcgt 1418

Db 1405 TCTCCAGCTGGCTGCTCTGCTGATGGAGCGCATGATCTCTGCTGCTGCTGCTCGT 1464

Qy 1419 ggtgaactctcacagcagcagtgagcagctgctcctctctctgagacacagcagtgag 1478

Db 1465 GGTGAACCTCTCACAGCCAGCAGGCTGCTCCTCTCTCTGAGACACGATGAGGTGAG 1524

Qy 1479 gacttaacttcagtgagttggtcacgtagcatcagatattgtgcacagactgatttgc 1538

Db 1525 GACTTACTTCATGAATTTGGTCACGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1584

Qy 1539 acgatttagcggaaacaaatgtgaaactgactttagagtgccatcgcaaatgcttga 1598

Db 1585 ACGATTTAGTGGACAAATGTGNAACCTGACTTTGTAGAGGTGCCATCAAAATGCTTGA 1644

Qy 1599 aaattgggtggagcgtcagttccctcccaagattgtcgaacattataaagatggaag 1658

Db 1645 AAACCTGGTGGACACACTGATCCCTCCGAAGGCTGCAAAACATTATAAAGATGGAAG 1704

Qy 1659 cctattgacagagatcgtgtgaaacattgttctttagctgagtggtgcacacaggtct 1718

Db 1705 CCCTATTACAGATGATGCTGTGAAACCTTGTGCTTCTAGACTGCTCAACACAGTCT 1764

Qy 1719 tctgacctgcgcagatgttttgagcaagtgtgacgtctcttcatcaccacacatc 1778

Db 1765 CTTGACCTTACGCAGATGTTTGTGACAAAGTTGATCACTCTCCACACACATTC 1824

Qy 1779 gctggatgctgcaagtgaatatgcgaatactgctgcgaataattagagtgagctac 1838

Db 1825 GCTGGATGCTCAAGTGAATGCCAATACTGCACAGAAATTTAGTGTGTCAGCTAC 1884

Qy 1839 tccaggcacaataatgcccagctaccttggacatttggcagggggatcacatgcccata 1898

Db 1885 TCCAGGCACAAATATGCGACTACTTTTGGACACTTGGCAGGGGATGATGCGCCAATA 1944

Qy 1899 ttatggatatcttggagtgagttatttccatggatatgttttacagctgttttaaaa 1958

Db 1945 TTATGGATATCTTTGGAGTGAAGTCCTTTCCATGGAGCATGTTTATAGCTGTTTTAAAAA 2004

Qy 1959 agaaggataatgaatccagaggttggaatgaatacagaacctaactcctgaaacctgg 2018

Db 2005 AGAAGGCAATGAATCCAGAGGTGGAATGAATACAGAACTATCTCTGAACTCG 2064

Qy 2019 gggatctctggacggcatgacatgctcccaatttctgaaacgtgagccaaacaaaa 2078

Db 2065 GGGATCCTCGAGCGCATGACATGCTTCCAAAATTTCTTGAACGTGAGCCAAACCAAAA 2124

Qy 2079 agcgttctaagttagtagcctgcactgctcgtga 2115

Db 2125 AGCTTCTTATGATGAGCGCTGCTGCTCGTAA 2161

RESULT 9

LOCUS AB000175

DEFINITION Porcine mRNA for endopeptidase 24.16, complete cds.

ACCESSION AB000175

VERSION AB000175.1 GI:1783133

KEYWORDS endopeptidase 24.16 type M3.

SOURCE Sus scrofa Adult Liver cDNA to mRNA.

ORGANISM Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

1 (bases 1 to 2728)

Direct Submission

Submitted (27-DEC-1996) to the DDBJ/EMBL/GenBank databases.

Shigehisa Hirose, Tokyo Institute of Technology, Department of Biological Sciences; 4259 Nagatsuta-cho, Midori-ku, Yokohama, Kanagawa 226-8501, Japan (E-mail:shihiro@bio.titech.ac.jp,

REFERENCE

AUTHORS Kato,A., Sugiura,N., Saruta,Y., Hosoiiri,T., Yasue,H. and Hirose,S.

TITLE Targeting of endopeptidase 24.16 to different subcellular compartments by alternative promoter usage

JOURNAL J. Biol. Chem. 272 (24), 15313-15322 (1997)

MEDLINE 97326108

FEATURES

source

Tel:045-924-5736, Fax:045-924-5824)

2 (sites)

Location/Qualifiers

1. .2728

/organism="Sus scrofa"

/db\_xref="taxon:9823"

/dev\_stage="Adult"

/tissue\_type="Liver"

1. .2728

/note="corresponding to exon3,4,5-16 of this gene;

mRNA

endopeptidase 24.16 type 3"

CDS

240. .2285

/standard\_name="endopeptidase 24.16"

/note="oligopeptidase M :neurolysin :SBAP(soluble angiotensin-binding protein) :MEP(microsomal metalloendopeptidase)"

/codon\_start=1

/product="endopeptidase 24.16 type M3"

/protein\_id="BAA19067.1"

/db\_xref="GI:1783134"

/translation="MTIGREAMSPLOAMSSYTVDRNLRWLSDPSQIKRRTEBELIAQ

TKOYDDIGMDIEEVTYENCLOALADVEVKYIVERTMLDFPOHVSDDKEXVRAATEA

QRLSRFDLEMSREDIFLRVLKETCDLGIKPEARYLLEKSVKMGKNGHLHPQ

VQNEIKAMKRMSELICIDFNKNUNEDDTFLVFSKAEALGALPDDFISLEKTDNNKYI

TLKYPHYFPMVKCCIPETRRKMEAFNTRKRENTIIQELLEPLRAKVLGLYSTH

ADPVELEMTAKSTHHVTAFLDLSOKLPLGEAREFILLNKKKEEKEFGEDGKIN

AWDLHYMTQTEELKYSVDEILYFPFIEVTEGLLNIYQELIGLSFEQVTDHVMN

KSVTLATVKDKATGEVLGOFYLDLYPREKYNHAACFGLOPCLLPDGRSMVSVAALV

VNFSQPRAGRPSLLRHDEVNTYEPHERGHVHQICAOIDFARFSGTNVTDFVEVPSOM

LENWWDTSRLRLSKHYKDGSPITDILLKELVASRLVNTGLLTLROIULSKVDQSLH

TNTSLDAASAYARKYTEILGVAATPGTNMPATFCHLAGGYDGYGYGLWSEVFSMDMF

YSCFKKEGIMNPEVMKRYNLILKPGSLDGMQLONFLKREPQNKAFLMSRLHAP"

BASE COUNT 847 a 518 c 616 g 747 t

ORIGIN

Query Match 86.4%; Score 1826.8; DB 4; Length 2728;

Best Local Similarity 92.3%; Pred. No. 0;

Matches 1921; Conservative 2; Mismatches 159; Indels 0; Gaps 0;

Qy 34 ctccgcagagttggtggtccaggattttactcaagaatgcgtttaggagaagaagtgtg 93

Db 204 CACATCTGGTGTGGTTCGGGATTTTATGAGAATGACTTTAGGAGAGAGAAGCAATG 263

Qy 94 tctcctctcaggaatgtcttctactgtgctgagcagaaatgttttaagtggat 153

Db 264 TCYCTCTTCAGGCAATGCTTCTCTACTGTGGATGGCAGAAATGTTTAAAGATGGAT 323

Qy 154 ctttaccagagcaaatataaacaagaactgagagctcatgttcgagaccacaacaggtg 213

Db 324 CTTTCACCAGACAAATCAAAAGAAAGAACTGAGGAGCTGATTCGCCAGACCAACAGGTG 383

Qy 214 taqatgctgttgaatgctcgttattgaggaagtaacttacgagaactgctcagagca 273

Db 384 TATGATGATATTGGAATGCTTGACATTTGAAGAGGTAACATTATGAGAACTGTTTCAGGCA 443

Qy 274 ctggcagatgtagaagtaagatatagtggaagagccatgctagactttccccagcat 333

Db 444 CTGGCAGATGTGAAGTGAAGTATATAGTGGACGACCACTAGACTTTCCCCAGCAT 503

Qy 334 gtatcctctgacaaagaagtacgagcagcagtaacagaagcagacacaaagactttctcgt 393

Db 504 GTCTCCTCTGACAAAGAGTACGAGCAGCAAGTACAGAAGCAGACAAAAGACTTTCCTGT 563

Qy 394 ttgatattgagatgagcagcagagagagatatatttggagaaattgttcattacagaaa 453

Db 564 TTTGATATTGAGATGAGCATGAGAGAAGATATCTTTCTGGAATGTTTCGTTTAAAGAG 623

QY 454 acctgtagctggggaagataaaacctgagccagacgatacttggaagcaatataa 513  
DB 624 ACCTGTGATTTGGGGAAGATCAAGCCTGAAGCAGAGATACCTAGAAAATCAGTAA 683  
QY 514 atggggaagaaatgggtccatctctctgaaacagatgaaatcaaatcaatg 573  
DB 684 ATGGGAAAAGGAATGGGTCCATCTTCGGGAACAAGTACGAATGAATCAAGCAATG 743  
QY 574 agaaagaatagtgagctatgtattgtatttaacaaaaacctcaatgagatatacc 633  
DB 744 AGAAAAGAATGAGTGAGCTATGTATGACTTCAACAAAACCTCAATGAGGACGATAC 803  
QY 634 ttcctgtattttccaggctgaacttggtctctctctctgattgatttgcagattta 693  
DB 804 TTCCTTGATTTTCCAAAGCTGAACCTGTGTCTTCCCTGATGATTTTCACACTTTA 863  
QY 694 gaaagacagatgacaaagtataaaatatacttaaaataccacactatttccctgctc 753  
DB 864 GAAAGACAGATGACAAAGATATAAAATTAACCTTAAAAATATCCACACTATTTTCCCTGTC 923  
QY 754 atgaagaatgttgtatccctgaaccacagaagatgaaatggcttttaatacaagg 813  
DB 924 ATGAAGAATGTGTATCTTGAACCAAGCAAGAAATGGAATGGCTTTAATACGAGG 983  
QY 814 tcaagagagaaacacaccataattttgcagcagctactcccactgcgaaccaaggtggcc 873  
DB 984 TGCAAGAGGAAAACACCATATCTTCGACGAGGTACTCCCACTGCGGGCCAAAGTAGCC 1043  
QY 874 aaactactggttatagacacatgctgactctgctctgaaatgaacactgcaaaagagc 933  
DB 1044 AAACACTCTGGGTACAGCACATGCTGACTTTGTCTTGAGATGAACACTCGGAAGAGT 1103  
QY 934 acaagccgtaacagcytttctagatgtttaagccagaagttaaaaccttggtgaa 993  
DB 1104 ACACACCATGTGACAGCCCTTTCTAGATGATTTAAGCCAGAAATTAACCCCTTGGGTGAG 1163  
QY 994 gcaaacgagagattttatttgatttgaaagaaaggaatgcaagacagaggggtttgaa 1053  
DB 1164 GCAGAACGAGAGTTTCAATTTGAATTTGAAGAAAAGGAATGTGAAGAGAAAGGTTTGA 1223  
QY 1054 tatgatggaaaataaatccctgggatactataattactacactacacagaggaactc 1113  
DB 1224 TATGATGGCAAAATCAATGCCCTGGATACATTAATATGACTCAGACAGAACAATC 1283  
QY 1114 aagatttccatgacaaagagttcctcaaggaatacttcccaatgaggtggtcactgaa 1173  
DB 1284 AGTACTCTGTGACCAAGAGATCTCAAGGAATACTTCCCAATTAAGTGTCTACTGAA 1343  
QY 1174 ggctgtgaaacacctaccagaggtgtgtgggactttcaattgaaacaaatgacagatgct 1233  
DB 1344 GCCTTGCTGAACATCTATCAGGAGTTGTGGGACTCTCCTTTGAGCAAGTGACAGATGCT 1403  
QY 1234 catgtttgaaacagagtggttacattttactgtgaaggataaagctacagagaagta 1293  
DB 1404 CATGTTTGAATAAAGTGTATACCTGTACACGGTGAAGGATAAAGCCACAGAGGAAGTA 1463  
QY 1294 ttggacagttctatttgacctctatccaagggaggaagaaatacaatcatcgccctgc 1353  
DB 1464 TTAGGACAGTTCTACTTGGACCTCTATCCAAGGAAGGAAAATACAAACCATCAGCCCTGC 1523  
QY 1354 ttccgtctccagcctggctgctctctgctgtggaagccgagatgagcagtggtgctgc 1413  
DB 1524 TTTGGTCTCCAGCCTGGCTCTCCTGCTGATGGAGCCGATGATGCTGTGCTGCTGCTGC 1583  
QY 1414 ctccgtgtgaaacttctcaagccagtggtgaggtgctccctctctctctctctctctctct 1473  
DB 1584 CTCGTGGTGAATCTTCACAGCCAGGAGCAGGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCT 1643  
QY 1474 gtgaggaacttctcatgagttggttcacgtgacatgacatgatttgcacagactgat 1533  
DB 1644 GTGAGACTTACTTCATGAATTTGGTCACGTCATGCAATGATTTGCGGACGACTGAC 1703  
QY 1534 tttgcacgatttagcggaacaaatgtggaacactgactttgttagaggtgacctgcgaatg 1593

DB 1704 TTTGCACGATTTAGTGGAAACAATGTGGAACATGACTTTGTAGAGTGCCATCACAAATG 1763  
QY 1594 cttgaaaattgggttgagcgtcattccctccgaagattgtcaaaacattataaagat 1653  
DB 1764 CTTGAAAATGGGTGTGGACACTGATTCCTCCGAAGCTGTCAAAACATTTATAAAGAT 1823  
QY 1654 ggaagccctattgcagacgactgcttgaaaaactgttgctcttaggctggtcacaaca 1713  
DB 1824 GAGGCCCTATTACAGATGATCTGCTTGAAAACCTTGTTGCTTCTAGACTGCTCAACACA 1883  
QY 1714 ggtcttcgacctggccagattgttttgagcaaaagttgacagttgacgtctctctcataccaa 1773  
DB 1884 GGTCTCCTGACCTTACGCCAGATTTGTTTGGAGCAAAAGTTGATCAGTCTCTCCACACCAAC 1943  
QY 1774 acatcgctgattgctgcaagtgaataatgcaaaactactgcaagaaatattagagattgca 1833  
DB 1944 ACTTCGCTGGATGCTGCAAGTGAATATGCCAAATCTGCACAGAAATTTTAGGTGTGCA 2003  
QY 1834 gctactccagggacaaaatgcccagctacttggacatttggcagggggatacagatgac 1893  
DB 2004 GCTACTCCAGGCACAAATATGCCAGCTACTTTTGGACACTTGGCAGGGGATATGATGCG 2063  
QY 1894 caatatttgatatctcttggtgagtgaaagtattttccatggatatgttttacagctggttt 1953  
DB 2064 CAATATTATGATATCTTTGGAGTGAAGTCTTTTCCATGGACATGTTTATAGCTGTTTT 2123  
QY 1954 aaaaagaagggataatgaatccagaggttggaatgaataacagaaaacatacctcgtgaaa 2013  
DB 2124 AAAAAAGAGGGATTAATGAATCCAGAGTTGGAATGAATAACAGAAACCTTAATCTGAAA 2183  
QY 2014 cctggggtactctggacggcatgacatgctcccaatttctgaaacgtgagcacaac 2073  
DB 2184 CTGGGGATCTCCTGGACGGCATGGACATGCTCCAAAATTTCTTGAACGTGAGCCAAAC 2243  
QY 2074 caaaagcgttccaatgagtagagcctgcatgctgcgtga 2115  
DB 2244 CAAAAGCGTTCCTAATGATGATAGAGGCGCTGCATGCTCCGTAA 2285  
  
RESULT 10  
RABEPD RABEPD 3507 bp mRNA MAM 03-FEB-1999  
LOCUS Rabbit mRNA for endopeptidase, complete cds.  
DEFINITION D13310  
ACCESSION D13310.1 GI:391848  
VERSION Zinc metalloendopeptidase; endopeptidase.  
KEYWORDS Oryctolagus sp. liver cDNA to mRNA.  
SOURCE Oryctolagus sp.  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
  
REFERENCE 1 (sites)  
AUTHORS Hrycyna,C.A. and Clarke,S.  
TITLE Purification and characterization of a novel metalloendopeptidase from Saccharomyces cerevisiae  
JOURNAL Biochemistry 32 (42), 11293-11301 (1993)  
MEDLINE 94032245  
REFERENCE 2 (bases 1 to 3507)  
AUTHORS Kawabata,S., Nakagawa,K., Muta,T., Iwanaga,S. and Davie,E.W.  
TITLE Rabbit liver microsomal endopeptidase with substrate specificity for processing proproteins is structurally related to rat testes metalloendopeptidase 24.15  
JOURNAL J. Biol. Chem. 268 (17), 12498-12503 (1993)  
MEDLINE 93286083  
COMMENT Submitted (28-SEP-1992) to DDBJ by:  
Shun-ichiro Kawabata  
Dept. of Biology  
Faculty of Science  
Kyushu University  
6-10-1 Hakozaki, Higashi-ku  
Fukuoka 812  
Japan  
Phone: 092-632-2742

FEATURES  
source

092-632-2742.  
Location/Qualifiers  
1..3507  
/organism="Oryctolagus sp."  
/db\_xref="taxon:60556"  
/tissue\_type="liver"  
154..2268  
/codon\_start=1  
/product="endopeptidase"  
/db\_xref="GI:391849"  
/translation="MIARCFSAVRGLHRVSRILFKMTLREVMSPQLQAVSVTAAG  
RNLRLWDSPOIKTRTEELIAOTKQVDSVGMLDIKDVTYENCLQALADAEVYKIVE  
RTMLDEPOHVSITREVRRAASTADKRLSRFDIESMRREDIFORLVHLOETCDLEKIKP  
EARYLSESVKGRNGLHLPPEVQNEIKSMKKRMSELCIDFNKNLNEDDTFLVFSKA  
ELGALPDFIDSLRMDDDKYITLKYPHYPPVMKKCCIPETRRMENAFNRKEEN  
TVIIQQPLPLRAVAKLIGYSTHADFLVLEMTAKTSRVTAFLLDDLSOKLPLGKEER  
EFILSKKEEKEEFYDGHVNPVSVTLTYTKDRATGEVLGDFVLDLYPRGKHVHQA  
LLNIYQELGLSPQVADAHVNPVSVTLTYTKDRATGEVLGDFVLDLYPRGKHVHQA  
CFGLQPCGLPDGSRMLSVAAVLVNFQPVAGRPSLLRLSKHYKDKGNPIADLLEKLIVAS  
QTFARSGTNVEDFVEVPSQMLENNWWDIDSLRLSKHYKDKGNPIADLLEKLIVAS  
RLVNTGLTLRQIVLSKVQDSQSHNTSSLDAASEIARYCTDILGVAATPGTNMPATFGH  
LAGYDGOYGYGLWSEVPSMDMFYSCFKKEGIMNPEVGMKYNLILRPGSLDGMML  
QNFLQREPNQKAFILMSRLQAP"  
3412..3417  
3416..3421  
3420..3425  
BASE COUNT 939 a 794 g 874 g 900 t  
ORIGIN

Query Match 85.1%; Score 1799; DB 4; Length 3507;  
Best Local Similarity 90.6%; Pred. No. 0;  
Matches 1916; Conservative 2; Mismatches 197; Indels 0; Gaps 0;  
QY 1 atgacgcgcgcgtgctttgtgctgtggaagcctccgcagagtggtggtccaggatt 60  
DB 154 ATGATCGCCCGTGTCTTTCGGCTGTGCGAGGCTCCACAGAGGTGGTGTCCAGGATT 213  
QY 61 ttactcagaatgacgttaggaagaagtgtatgtctctctctcagcgaatgtctctctat 120  
DB 214 CTATTCAAAATGACATTTAGGAGAGAAAGTGATGTCACCTCTTTCAGCGCAGTTTCTTCTAT 273  
QY 121 actgtggctggcagaaatgttttaagatgggactttccacagagcaaatataaacaaga 180  
DB 274 ACTGCAGCTGGCAGGAGTGTATAGATGGGATCTTTCACAGAGCAAAATCAAGACAGA 333  
QY 181 actgaggagctcattgtgcagaccacacagggtgtacgatgctgttggaaatgctcgatt 240  
DB 334 ACTGAGGAGCTCATTGCGCAGACACAGCAGGTTGATGATTCCTGTTGGAATGCTGGATATC 393  
QY 241 gaggaaagtaactacagaaactgtctgaagcactggcagatgtagaadtgaagtata 300  
DB 394 AAGGACGTGACTTACGAGAACTGCTTTCAGGCGCTGGCAGATGTGGAAGTGAAGTACATA 453  
QY 301 gtggaagaccatgctgaactttccacagcatgtatcctctgacaaagaagtacagca 360  
DB 454 GTGGAAGAACCATGCTAGACTTCTCAGCATGTTTCTACATGACAGAGAAGTACGGCA 513  
QY 361 gcaagtacagaagcagacaaaagacttctctgttttattgattgagatgagcatgagaga 420  
DB 514 GCAAGTACAGAAGCGACAAAGAGGCTTCTCGCTTTTGATATTGATGATGATGAGAGAA 573  
QY 421 gatatattgagagaattgttcaattacagaaacactgtgactcgtgggaagataaaacct 480  
DB 574 GATATATTTTACAGAAATCGTTTCACTTACAGAAACCTGTGATCTGGAGAGATAAAGCCG 633  
QY 481 gaggccacgactactgaaagaagcaaatataatgggaaagaagaatggcctcattct 540  
DB 634 GAAGCCAGACGATCTGTGAAAGTACAGTTAAATTTGGGAAGAGGAATGGGTCCACCTT 693  
QY 541 cctgacaagtagcagaatgaaatcaaatcaatgaagaaagaatgagtgactgtatt 600  
DB 1774 TCTCTCCGAAGATTATCAAAACATTTAAAGATGGAACCCCTATTTCAGATGATCTTCTT 1833

DB 694 CCAGAAGAAGTACAGAAATGAAATCAATCAATCAAGAAAAAATGAGTGAGCTCTGTATT 753  
QY 601 gattttaacaaaaaacctcaatgagatgatataccttctctgttatatttccaggctgaactt 660  
DB 754 GACTTTTAACAAAAACCTCAATGAGGATGATACCTTCTTGTGTCTTTTCCAAGGCTGAGCTT 813  
QY 661 ggtgctcttcctgatgatttcttgcacagttagaaaaagacagatgatgacaagtataaa 720  
DB 814 GGTGCTCTTCCGTCGATGATTTCAATGACAGTGTAGAAAAAGATGATGATCAAGATATAA 873  
QY 721 attactttaaaatataccacactatttccctgtcatgaagaaatgttgtatccctgaaacc 780  
DB 874 ATTACTTTAAAAATATCCACACTATTTTCTGTGTCATGAAGAAATGTTGTATTCTCTGAACC 933  
QY 781 aagaagaagatggaatggcttttaatacaaggtgcaaaagagagaaaaaccataatttgg 840  
DB 934 AGAAGAGAGAAATGGAATGGCTTTTAAATACAAGATGCAAGAGAGAAAAACCGTCATCTG 993  
QY 841 cagcagctactccactcgaaaccaggtggcgaactcgtggttataagcacacatgt 900  
DB 994 CAGCAGCTACTCCACTGGGGGCCCAAGTGGCCAAACTCTCTGGCTATAGCACACATGCT 1053  
QY 901 gacttcgtcttgaatgaactgcaaaagagacaaagcgcgcgttaacagcyytttcttagat 960  
DB 1054 GACTTCGTCTCGAGATGAACACTGCAAGAGACACAAGCCGAGTAACAGCCTTTCTTAGAC 1113  
QY 961 gatttaagccagaagttaaaacccctgggtgaagcagaagagaggtttattttgaatttg 1020  
DB 1114 GATTTAAGCCAGAAATTAACACCGCTGGGTGAGGCAAGAACGAGAGTTTATCTTGAGTTTG 1173  
QY 1021 aaaaaaagaatgcaaaagcagagggttttgaatatgtatgggaaaaatcaatgctcgtggat 1080  
DB 1174 AAAAAAAGAAATGCAAGAGAAAGGTTTTGAGTATGACGGGAAAAATCAATGCTCGGGAC 1233  
QY 1081 ctattactacatgactcagacagagggaactcaagtattccatagacaaagagtttccctc 1140  
DB 1234 CTGCATTACTACATGACTCAGACAGAGAGACTCAAGTACTCCATCGACCAGGAGTTTCATC 1293  
QY 1141 aaggaataacttcccaattgaggtggtcactgaaggtctgtctgaacacactaccaggagttg 1200  
DB 1294 AAGGAGTACTTCCGATTGAGGTGCTACCGAGGGGCTGCTGAACATCTACCAAGGAGTTG 1353  
QY 1201 ttggacttctcatttgaacaaatgacagatgctcattgttggaaacagagtggtacactt 1260  
DB 1354 CTGGGACTTTTCATTCAGCAAGTGGCCGACGCTCATGTTTGGNACCCGAGTGTCACACTT 1413  
QY 1261 tatactgtgaagataaaactacagagaagatattgggacagttctatttggacctctat 1320  
DB 1414 TACACTGTGAAGGACAAGGCTACAGGAGAAGTGTCTGGGACAGTTCTACCTGGACCTCTAT 1473  
QY 1321 ccaaggaaggaataatacaatcatcgccgtcgtcgtctccagcgtgctgcctctctg 1380  
DB 1474 CCGAGGAAGGAAAAATATAATCATGCAGCTGCTTGTGTCTCCAGCTGGCTGCCCTGCTG 1533  
QY 1381 cctgtggaagcggagatgatggcagtggtcctcgtgtgtgaacttctcaagcagagtg 1440  
DB 1534 CCGGATGGGAGCGGATGCTGCTGTGGCTGCCCTCGTGTGTAACCTTCTCCAGCCAGTG 1593  
QY 1441 gcaggtcgttcctctctctcctgagacacagcagaggtgaggaacttaacttaagtttgg 1500  
DB 1594 GCCGGCGGGCCCTCTCTGCTGAGGCACGATGAAGTGAGGACTTACTTCGATGAGTTTGGT 1653  
QY 1501 caggtgatcatcagatttggcacagactgatttgcagattttagcggaacaaatggtg 1560  
DB 1654 CATGTCATGCATCAGATTTGTGCACAGACTGATTTTTCACGATTTTAGTGAACAAATGTG 1713  
QY 1561 gaaactgactttgtagagtgccatcgcaaatgttgaataatgggtgtggagcgtcgat 1620  
DB 1714 GAAACTGACTTTGTAGAGGTGCTTTCACAAATGCTTTCAGAACTGGGTGTGGGACATCGAC 1773  
QY 1621 tccctccgaagattgtcaaaacattataaagatgggaagccctatttgcagacatcgtt 1680  
DB 1774 TCTCTCCGAAGATTATCAAAACATTTAAAGATGGAACCCCTATTTCAGATGATCTTCTT 1833



Qy	721	attacctaaatatccaaactatttcccctgctcatgaagaatgttgtatccctgaacc	780
Db	864	GTTACCTAAAGTATGCTCACTACTTCCCTGTGATGAAGAAATGCTGTCCCTGAAACT	923
Qy	781	agaagaaggaatgaatggcttttaatacaaggtgcgaagaggaaaaacaccataatttg	840
Db	924	CGAAGGAAGATGAATGGCTTTTCATACACGAGTGCAAACAGGAACACCGCAATTTCTG	983
Qy	841	cagcagctactcccactgcgaaccaagggtggccaaactactcggttatagcacacagtct	900
Db	984	CAACAGCTCTCCCGCTGGGAGCTCAAGTGGCCCAAGCTCTGGGCTATAAACACCCATGCT	1043
Qy	901	gattctgctctgaatgaacactgcgaagagacgaacgcgcgttaacagcytttctagat	960
Db	1044	GACTTTTCTCTTGAACATGAACACTCGGAAGAGACACAGCCGCGTGGCGCTCTTCTAGAT	1103
Qy	961	gatttaagccagaagttaaacaccttgggtgaagcagaacagagatttatttgaatttg	1020
Db	1104	GATTTAAGCCAGAAATTGAACCCCTTGGGTGAGGCAGAACGGGAGTTTATTTTGAAGCTG	1163
Qy	1021	aagaaaaggaatgcgaagacaggggttttgaatatgatgggaaaaatacaatgcctgggat	1080
Db	1164	AAGAAGAAGGAATGTGAAGAGCGAGGCTTTTGAATACGACGGGAAGATCAATGTCATGGGAC	1223
Qy	1081	ctatatctactgactcagacagaggaagaaactcaagtattccatagacacaagagttcctc	1140
Db	1224	CTCCATTTACTTACATGACACAGACAGAGGAGCTCAAGTACTCGGTGGACAGGAGTCGCTT	1283
Qy	1141	aaggaatacttcccattgaagttggttcactgaagcttgcctgaacactaccagagttg	1200
Db	1284	AAGGAGTACTTCCCATTTGAGTGGTTCAGGAAGGCTTGCTTAAGCATCTACAGGAGCTG	1343
Qy	1201	tgggactttcatttgaaacaaatgcagatgctcatgttttgaaacaagagtgtcacactt	1260
Db	1344	CTGGGACTTTCATTTGAACAAAGTCCCGATGCTCATGTTTGGAAATAGAGGCTTTCATT	1403
Qy	1261	tatactggaagataaagctacagagagaagtbatyggacagttcttatttggacctctat	1320
Db	1404	TACACTGTGAAGGATAAAGCTACAGAGAGAGTCTCTGGGGCAGTTCTACTCTGGACCTCTAT	1463
Qy	1321	ccaaggaaggaataataatcatcgccctgcttcgcttcagctccagcctggctgctcttg	1380
Db	1464	CCAAGGAAGAAATATACAACTGAGGCTGCTTCGGTCTCCAGCGAGGCTGCTTCTC	1523
Qy	1381	cctgatggaagccggatgatggcagtgggctgccttcgcttggtgaactctctcacagccagt	1440
Db	1524	CCTGATGGCAGTCGGATGATGCTGTGGCGCCGCTTGCTGCTCTCTCAGCCGTGA	1583
Qy	1441	gcaggtgcctctctctctgagaacgcagcagggtaggaacttacttcatagatttgggt	1500
Db	1584	GCAGGTGGCCCTCTCTCTGAGACATGATGAAGTGGCGACTTACTTTCACGAGTTGCGT	1643
Qy	1501	cacgtgatcatcagattgtgcacagactgatttgacagatttagcggaacaaatgtg	1560
Db	1644	CACGTATGATCAGATCTGTGGCAGACTGACTTTGACCATTCAGTGGGAACAAACAGTG	1703
Qy	1561	gaaactgactttgtagaggtgccactgcgaatagcttgaaatttgggtgtggagcgtcgat	1620
Db	1704	GAACTGACTTTGTAGAGGTGCCATCACAATATGCTTGAANAACCTGGGTGTGGAGAGTTGAT	1763
Qy	1621	tccctccgaagattgtcaaaacattataaagatggaaagccctattgcagacgactctgctt	1680
Db	1764	TCCCTGGCAAACTGTCAAAAGCATTTATAAGATGGACACCCCTATCAGACGAGGCTGCTC	1823
Qy	1681	gaaaaacttgtcttctaggtcgttcaacacaggttcttgcaccttcgcgcagatttgtt	1740
Db	1824	GAGAAGCTTGTGTCTTAGCTTGGTGGTCAACACAGGCTTCTCTGACGCTGGCGCAATTTGT	1883
Qy	1741	ttgacgaagtgtgatcagctctctcatccacaacatcgctggatgctgcgaagtgaatat	1800
Db	1884	TTGAGCAAAAGTTGACCAATCTCTCCATACCAATCGACTCTGGATGGCGGAGGGAATAC	1943
Qy	1801	gccaaatactgctcagaataatttaggagttgcagctactctccagcgaacaaatlatcgcaqct	1860

Db	1944	GCTAAATACTGCAGAAATTC	TGGGAGTGTGAGTACGCCAGGCACAAATATGCCAGCT	2003
Qy	1861	acctttgacatttgcaggggatcacgatgcccaattatttgatatcttttgagtgaa	1920	
Db	2004	ACTTTTGGACACTTGGCAGGAGGGTATGATGCCAAATATATATGGATATCTTTGGAGTGAA	2063	
Qy	1921	gtatttccatgatatgtttttacagctgttttataaaagaaggataatgaatccagag	1980	
Db	2064	GTGTTTTCCATGGACATGTTTTCACACCTGTTTTAAAAAGAGGGGATCATGAATCCAGAG	2123	
Qy	1981	gttgaatgaatacacgaaacctaacttcctgaacacctggggatctctctggacggcatggac	2040	
Db	2124	GTTGGAATGAATATACAGAACTTTAAATCCTCAAGCCTGGGGGTCCCTCGAGCGCATGGAC	2183	
Qy	2041	atgctccacaatttcttgaaacgtgagcaaaccaaaagcgttctctaagttagaggc	2100	
Db	2184	ATGCTCCAGAAATTTCTTGCAACGTGAGCCAAACCAAAAGCATTCCTAATGAGTCGAGGC	2243	
Qy	2101	ctgcatgctcgtga	2115	
Db	2244	CTGAATGGTTCGTAA	2258	
RESULT 12				
BC004985				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostom				
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
1 (bases 1 to 2637)				
Direct Submission				
Submitted (21-MAR-2001) National Institutes of Health, Mammalia				
Gene Collection (MGC), Cancer Genomics Office, National Cancer				
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590				
USA				
NIH-MGC Project URL: http://mgc.nci.nih.gov				
Contact: MGC help desk				
Email: cgabbs-remail.nih.gov				
Tissue Procurement: DCTD/DTp				
cDNA Library Preparation: Rubin Laboratory				
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)				
DNA Sequencing by: Genome Sequence Centre,				
BC Cancer Agency, Vancouver, BC, Canada				
info@bcgsc.bc.ca				
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,				
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin				
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, So				
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven				
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saedi, Jacques				
Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Sto				
Michael Thorne, Miranada Tsai, Natasha van den Bosch, Jill Vard				
George Yang, Scott Zuyderduyn, Marco Marra.				
Clone distribution: MGC clone distribution information can be f				
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov				
Series: IRAL plate: 9 Row: p Column: 22.				
Location/Qualifiers				
1. .2637				
/organism="Homo sapiens"				
/db_xref="taxon:9606"				
/clone="MGC:3978 IMAGE:2958628"				
/tissue.type="kidney, renal cell adenocarcinoma"				
/clone_lib="NIH_MGC_14"				
/lab_host="DHI08-R"				
/note="Vector: pOTB7"				
117. .1940				
CDs				

```
/codon_start=1
/product="Unknown (protein for MGC:3978)"
/protein_id="AA04985.1"
/db_xref="GI:13436419"
/translation="MIRALCLVARSRLRVGSRILLRMTLGRVNSPLQAMSVTVAG
RNVRLDLSPEQIKRTVEELIVQKVYDAVMGLIEEYTYENCLOALADVEVKYIVE
RTMLDPPOHVSDEKRAASTADKRLSRPDIEMSRGDIENFVHLQEDCDLQKIKP
EARVLEKSIKMGKRNGLHLPQVONETKSMKRMSELCIDFNKLNEDDFTLVFSA
ELGALPHDFIDSLKTDKDKYIKLKYPHYFVPMKKCIPETRRMEMAFNTRKEEN
TIILQOLPLPLKTVAKLGLVSTHADFLVEMNTAKSTSRVTAFLDDLQSLKPLGEAER
EFILNKKKECKDFEYDGIINAWDLYMTQTEELKYSIDQFLKYEYFIEVYTEG
LLNTYELLGLSPQMDAHWNKSVTLTVKDKATGEVLQDFYLDLYPREGKNHAA
CFNGLOPCLLPDGMMAVAALVFNOPVAGRPSLLRHDEVRYEFHEFGVHQICA
QVSFFPVPNLPISEFKLLDCCOVSAPNRFQKLVNFKNFILGLLQNLQNPYPVLFHN
IFLGGSTSERCLNLSLWKNKTKTIERQLICFFLLLL"
BASE COUNT 790 a 522 c 592 g 733 t
ORIGIN

Query Match 72.2%; Score 1526; DB 9; Length 2637;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1529; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 atgacgcgcgcgtgctcttggctgtgcgaagcctccgcagagttgggtgtccaggatt 60
DB 117 atgacgcgcgcgtgctcttggctgtgcgaagcctccgcagagttgggtgtccaggatt 176

QY 61 ttactcagaatgacctaggagaagaatgtagtctctctctcctcaggaatgtctcctcat 120
DB 177 ttactcagaatgacctaggagaagaatgtagtctctctcctcaggaatgtctcctcat 236

QY 121 actgtgctgcagaaatgttttaagatggatctctcaccagacaataataaacaaga 180
DB 237 actgtgctgcagaaatgttttaagatggatctctcaccagacaataataaacaaga 296

QY 181 actgaggagctattgtgcagaccacaaggtgtacgatgctgttgggaatgctcggtatt 240
DB 297 actgaggagctattgtgcagaccacaaggtgtacgatgctgttgggaatgctcggtatt 356

QY 241 gaggagaacttacagaactgtctgcagcactggcagatgtagaagtaaaagtata 300
DB 357 gaggagaacttacagaactgtctgcagcactggcagatgtagaagtaaaagtata 416

QY 301 gtggaagaccatgactgacttcccgacatgtatcctctgacaaagaagtcagaca 360
DB 417 gtggaagaccatgactgacttcccgacatgtatcctctgacaaagaagtcagaca 476

QY 361 gcaagtacagaagcagacaaagactttctcgttttgatattgagatgagcatgagaga 420
DB 477 gcaagtacagaagcagacaaagactttctcgttttgatattgagatgagcatgagaga 536

QY 421 gatatattgagagaattgttcattacaggaacactgtgatctctggggaagataaaacct 480
DB 537 gatatattgagagaattgttcattacaggaacactgtgatctctggggaagataaaacct 596

QY 481 gaggccagacatacttgaaaagtcaattaaaatgggaaaagaatgggctccatctt 540
DB 597 gaggccagacatacttgaaaagtcaattaaaatgggaaaagaatgggctccatctt 656

QY 541 cctgacaagtcagaatgaatacaaatcaatgaagaaaagaatgagtgagctatgtatt 600
DB 657 cctgacaagtcagaatgaatacaaatcaatgaagaaaagaatgagtgagctatgtatt 716

QY 601 gattttaaaaaaacctcaatgagatgataacctcctctgtatttttccaaaggtgaactt 660
DB 717 gattttaaaaaaacctcaatgagatgataacctcctctgtatttttccaaaggtgaactt 776

QY 661 ggtgctctcctgatgtattcattgacagtttagaaaagacagatgatacaagtataaa 720
DB 777 ggtgctctcctgatgtattcattgacagtttagaaaagacagatgatacaagtataaa 836

QY 721 attaccttaaatatccacacatttccctgtcatgaagaagaatgttgtatccctgaaacc 780
```

```
DB 837 ATTACCTTAAATAATCCACACTATTTCCTGTGTCATGAAGAAATGTTGTATCCCTGAAACC 896
QY 781 aqaagaagatgagaatggcttttaatacaaggtcaaaagagagaaacacacataatttg 840
DB 897 AGAAGAAGGATGGAATGCTTTTAAATCAAGGTCAAAGAGGAAGAAACACCAATAATTGG 956
QY 841 caqagctactccccactgcgaaccaaggTggccaaactactcggttatagcacacatgct 900
DB 957 CAGCAGCTACTCCCACTGGGAACCAAGGTGGCGCAAACTACTCGTGTATAGCACACATGCT 1016
QY 901 gacttcgctcttgaatgaacactgcaagagacacaaagccgcgcgttaacagcyttttctagat 960
DB 1017 GACTTCGTCCTTTGAAATGAACACTGCAAGAGACACAAAGCCGCTAACAGCCTTTCTAGAT 1076
QY 961 gatttaagccagaagttaaaaccccttgggtgaagcagaacagagagtttatatttgaaattg 1020
DB 1077 GATTTAAGCCAGAAAGTTAAACCCCTTGGTGAAGCAGACGAGAGTTTATTTTGAATTTG 1136
QY 1021 aagaaaaaggaatgcaaaagacaggggttttgaatatgatgggaaaaatcaatgcctgggat 1080
DB 1137 AAGAAAAAGGAATGCAAGACAGAGGGGTTTGAATATGATGGGAAAAATCAATGCCCTGGGAT 1196
QY 1081 ctatatctactacatgactcagacagacagaaactcaagtattccatagaccaagagttccctc 1140
DB 1197 CTATATTACTACATGACTCAGACAGAGAACTCAAGTATTCCATAGACCAAGAGTTCCCTC 1256
QY 1141 aaggaatacttcccgaattgaggtggtcactgaaggtctgtgaacacactaccagaggttg 1200
DB 1257 AAGGAATACTTCCCAATTGAGGTGCTCACTGAAGGCTTGCTGAACACCTTACCAGGAGTTG 1316
QY 1201 ttgggacttctcaattgaaacaaatgacagatgctcattgttggaaacaaagtggttacactt 1260
DB 1317 TTGGGACTTTCATTTTGACAAATGACATGCTCATGTTTGGAAACAAGTGTTACACTT 1376
QY 1261 tatactgtaaggataaaagctacaggaagaagtattgggacagttctatttggacctctat 1320
DB 1377 TATACTGTGAAGATAAAGTACAGAGAAAGTATTGGGACAGTTCTATTTTGGACCTCTAT 1436
QY 1321 ccaaggggaagaaaaatacaataatcgcgccctgctcgtgctccagcctggtgcctcttg 1380
DB 1437 CCAAGGGAAGGAAAAATACAATCATCGGCTGCTTTCGGGTCTCCAGCCTGGCTGGCTTCG 1496
QY 1381 cctgtggaagcggatgatggcagtgctgcctcctcggtgaactcttcacagccagtg 1440
DB 1497 CCTGATGAAGCCGGATGATGCGCAGTGCTGCCCTCGTGGTGAACCTTCTCAGACGCAATG 1556
QY 1441 gaagtcctcctctcctcctgagacacagcagagtgagagacttacttcaatgagtttgtt 1500
DB 1557 GCAGGTCGCTCCTCTCTCTGAGACACGACGAGGTGAGGACTTACTTTTCATGAGTTTGGT 1616
QY 1501 cactgtgatgcacagatttggcacagactgatttt 1536
DB 1617 CACGTGATGCATCAGATTGTGTCACAGAGGTGAGTTT 1652
```

```
RESULT 13
BC001644
LOCUS BC001644 2646 bp mRNA PRI 12-JUL-2001
DEFINITION Homo sapiens, Similar to thimet oligopeptidase 1, clone MGC:2361
IMAGE:2958628, mRNA, complete cds.
ACCESSION BC001644
VERSION BC001644.1 GI:12804472
KEYWORDS MGC..
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2646)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (08-JAN-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
```





Db 1566 GCAGGTGCTCCTCTCCTGAGACACGACGAGGTGAGGACTTACTTTCATGAGTTGGT 1625  
Qy 1501 cactgtagcatcagatttgcacagactgatttt 1536  
Db 1626 CACGTGATGCATCAGATTGTGCACAGTGAGTTT 1661  
RESULT 14  
AB033052 AB033052 5343 bp mRNA PRI 11-NOV-1999  
DEFINITION Homo sapiens mRNA for KIAA1226 protein, partial cds.  
ACCESSION AB033052  
VERSION AB033052.1 GI:6330675  
KEYWORDS  
SOURCE 'Homo sapiens brain cdna to mRNA, clone\_lib:pBluescriptII SK plus  
clone:fh04414.'  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (sites)  
Nagase,T., Ishikawa,K., Kikuno,R., Hirose,M., Nomura,N. and  
Ohara,O.  
TITLE Prediction of the coding sequences of unidentified human genes. XV.  
The complete sequences of 100 new cDNA clones from brain which code  
for large proteins in vitro.  
(DNA Res. 6 (5), 337-345 (1999))  
JOURNAL 20039619  
MEDLINE 2 (bases 1 to 5343)  
AUTHORS Ohara,O., Nagase,T. and Kikuno,R.  
TITLE Direct Submission  
JOURNAL Submitted (04-OCT-1999) to the DDBJ/EMBL/GenBank databases. Osamu  
Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology;  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail:cdnaif@kazusa.or.jp, URL:http://www.kazusa.or.jp/huge/,  
Tel:+81-438-52-3913, Fax:+81-438-52-3914)  
FEATURES  
source  
1. 5343  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="pbluescriptII SK plus"  
/tissue\_type="brain"  
1. 1345  
/gene="KIAA1226"  
1. 1345  
/gene="KIAA1226"  
/note="Start codon is not identified."  
/codon\_start=2  
/product="KIAA1226 protein"  
/protein\_id="BAA86540.1"  
/db\_xref="GI:6330676"  
/translation="PETRRREWAFNTRCKEENTIILOQLPLRTKVAKLLGYSTHAD  
FVLEMTAKSTSRVTAFLDLSQLKPLGEAREFILNLKKKECKDRGPEYDGKINAW  
DLIYYMTBELKYSIDQELKEFYFIEVYVTEGNTQYBELGLSPFQMTDAHWNKS  
VTLTYVKDKATGVELGYFDLPLPREKYNHAAFCFGLQPCLLPDGSRMMAVALVYN  
FSQPVAGRSELLRHDVRYFHEFGVHMHCIAQTDFAFSGTNVDFDFEVSQMLE  
NWYWDVLSRLSKHYKDGSPIDDLLEKLVASRLVNTGLTLROI VLSKVDOSLHTN  
TSLDAASEYAKYCEILGVAAVTPCTNMPATFGLAGYDQGYGLWSEVFSMDMEYS  
CFKKEGIMNFVEVMKYNLILKPGGSLDGMHLNLFUKRPFNOKAFLMSRGLHAP"  
BASE COUNT 1607 a 1060 c 1107 g 1369 t  
ORIGIN  
Query Match 63.68; Score 1344.2; DB 9; Length 5343;  
Best Local Similarity 99.99; Pred. No. 0;  
Matches i1343; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Qy 771 ccctgaaaccagaagagatggaattgcttttaatacagaagtgcaaggaggaacac 830  
Db 1 CCCTGAACACAGAGAGAGGATGGAAATGGCTTTAATACAGAGGTGCAGAGGAAACAC 60  
Qy 831 cataatttgcagcagctactcccactgcgaaccaaggtggccaactactcgttatag 890

Db 61 CATAATTTTGAGCAGCTACTCCCCTGCGAACCAAGGTGGCCAAACTACTCGGTATAG 120  
Qy 891 cacacatgctgaacttcgtcccttgaaatgaacactgcaagaagcacaagccggttaacagc 950  
Db 121 CACACATGCTGACTTCGTCCTTGAATGAACACTGCAAGAGACCAAGCCGCGTAACAGC 180  
Qy 951 yttctagatgatttaagccagaagttaaacccttgggtgaagcagacagagagtttat 1010  
Db 181 CTTTCTAGATGATTAAGCCAGAGTTAAACCCCTTGGGTGAGCAGACAGAGCTTAT 240  
Qy 1011 ttgaaatttgaaaaaagggaatgcaaaagacaggggtttttgaatatgatggaaaaatcaa 1070  
Db 241 TTTGAATTTGAAGAAAAAGGAATGCAAGACAGAGGGGTTTGAATATGATGGAAAAATCAA 300  
Qy 1071 tgcctgggatactatattactacatgactcagacagagagaactcaagtattccatagacca 1130  
Db 301 TGCCTGGGATCTATATTACTACTGACTCAGACAGAGGAACCTCAAGTATTCATAGACCA 360  
Qy 1131 agagtctctcaaggaatacttcccaattgagtggtggtcactgaaggtcttgcaaacaccta 1190  
Db 361 AGAGTTCTCTCAAGGAATACTTCCCAATTGAGGTGCTCACTGAAGGCTTGTGAACACCTA 420  
Qy 1191 ccaggagtgttgaggacttctcatttgaaacaaatgacatgctcactgtttggaaacaag 1250  
Db 421 CCAGGAGTTGTTGGGACTTTCATTTGAACAAATGACAGATGCTCATGTTTGGAAACAAG 480  
Qy 1251 tgttacactttactactgtaagataaagctacaggaagattattggacagttctatt 1310  
Db 481 TGTTACACTTATCTGTGAAGGATAAAGCTACAGGAGAAGTATTGGACAGACTTCTATT 540  
Qy 1311 ggacctctatccaaaggggaagaaatacaatcatctgcgcctgcttcgctccacagcctgg 1370  
Db 541 GGACCTCTATCCAAAGGGAAGAAAAATAACAATCATGCGCCTGCTTCGCTCTCCAGCCTGG 600  
Qy 1371 ctgcctctgcctgatggaagccggatgatggcagtggtcgtccctcgctgggaaactcttc 1430  
Db 601 CTGCTTCTGCTGATGGAAGCCGGATGATGCGAGTGTGCTGCTGCTGCTGCTGCTGCTC 660  
Qy 1431 acagccagtgcaggtcgtccctctctctctctctctctctctctctctctctctctctctca 1490  
Db 661 ACAGCCAGTGGCAGTGTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720  
Qy 1491 tgagtttgctcagctgtagcatcagattgtgacagactgatttgcacagactgatttgcacag 1550  
Db 721 TGAGTTTGTGTCAGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780  
Qy 1551 acaaaatgtaaaactgacttctgtagagtgccatcgcaaatgcttgaataatgggtgtg 1610  
Db 781 AACAAATGTGAAACTTGACTTTGTAGAGTGCCATCGCAAAATGCTTGAATAATTTGGGTGTG 840  
Qy 1611 ggagctgattccctccgaagatgtcaaaacattataaagatggaagccctcattgcaga 1670  
Db 841 GGAGCTGCAATCCCTCCGCAAGATGTCAAAACATTTAATAAGATGGAAGCCCTATTGCGAGA 900  
Qy 1671 cgatgctgtgaaaaaacttgttctcttagctggtgcacacaggtctcttgacacctgcg 1730  
Db 901 CGATCTGCTTCAAAAACCTTGTGCTTCTAGCTGTGTCAACACAGGTCTTCTGACCCCTGCG 960  
Qy 1731 ccagattgttttgagcaaatgtgatcagctctctctctctctctctctctctctctctctctc 1790  
Db 961 CCAGATTGTTTTGAGCAAAAGTTGATCAGTCTCTTCATACCAACACATCGCTGGATGCTGCT 1020  
Qy 1791 aagtgaaatgccaaatctactgctcagaaatattaggagttgcagactactccaggacacaaa 1850  
Db 1021 AAGTGAATATGCCAAATACTGCTCAGAAATATTAGGAGTTGACGCTACTCCAGGCACAAA 1080  
Qy 1851 tatccagctacacctttggacatttggcagggggatcacgatggccaatattatggatatct 1910  
Db 1081 TATGCCAGCTACCTTTGGACATTTGGCAGGGGGATACCATGGCCAATATTATGGATATCT 1140  
Qy 1911 ttgagtgaaagtatttccatggatatgttttaacagctgttttaaaaaaagagggataat 1970





Db 907 TGGCTGGATGCTGCAAGTGAATATGCCAAATACTGCTCAGAAATATTAGGAGTTCCAGCT 966  
QY 1837 actcaggcacaaatatccagctacctttggacatttggcaggggatacagatggccaa 1896  
Db 967 ACTCCAGGCACAAATATGCCAGCTACCTTTGGACATTTGGCAGGGGGATACGATGCCCAA 1026  
QY 1897 tattatgatatctttggagtgaaagtattttccatggatatgttttacagctgtttttaa 1956  
Db 1027 TATTATGGATATCTTTGGAGTGAAGTATTTCCATGGATATGTTTTACAGCTGTTTTAAA 1086  
QY 1957 aaagaaggataataatgaatccagaggttggaaatgaatacagaacctaatacctgaaacct 2016  
Db 1087 AAAGAAGGGATAAATGAATCCAGAGTTGGAATGAATACAGAAACCTAATCCTGAAACCT 1146  
QY 2017 ggggatctctggacgcatggacatgctccacaatttcttgaacgtgagccaaaccaa 2076  
Db 1147 GGGGATCTCTGGACGGCATGGACATGCTCCACAAATTTCTTGAACGTGAGCCAAACCAA 1206  
QY 2077 aaagcgttcctaatagtagtaggacctgcatgctycgtga 2115  
Db 1207 AAAGGTTCCCTAATGAGTAGAGGCCCTGCATGCTCGGTGA 1245

Search completed: January 14, 2002, 20:25:35  
Job time: 5144 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 18:56:01 ; Search time 106.63 seconds  
(without alignments)  
965.729 Million cell updates/sec

Title: US-09-833-782-2  
Perfect score: 3668  
Sequence: 1 MIARCLLAVSLRRVGGSR.....FLKRPNQAFAMSLRGLHAP 704

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL17.\*
- 1: sp.archaea.\*
  - 2: sp.bacteria.\*
  - 3: sp.fungi.\*
  - 4: sp.human.\*
  - 5: sp.invertebrate.\*
  - 6: sp.mammal.\*
  - 7: sp.mhc.\*
  - 8: sp.organelle.\*
  - 9: sp.phage.\*
  - 10: sp.plant.\*
  - 11: sp.rodent.\*
  - 12: sp.virus.\*
  - 13: sp.vertebrate.\*
  - 14: sp.unclassified.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3668	100.0	704	4 Q9BYT8	Q9BYT8 homo sapien
2	3441	93.8	745	6 P79433	P79433 sus scrofa
3	2640	72.0	607	4 Q9BQD0	Q9BQD0 homo sapien
4	2355	64.2	447	4 Q9ULJ4	Q9ULJ4 homo sapien
5	2299	62.7	687	11 Q9EPX1	Q9EPX1 mus musculu
6	2291.5	62.5	685	13 Q9PTV2	Q9PTV2 xenopus lae
7	1703	46.4	480	4 Q9BW75	Q9BW75 homo sapien
8	1177	32.1	716	3 P97996	P97996 aspergillus
9	1017.5	27.7	998	10 Q9FXD5	Q9FXD5 arabidopsis
10	848	23.1	680	2 Q9KVF8	Q9KVF8 vibrio chol
11	804	21.9	680	2 Q9LA87	Q9LA87 aeromonas h
12	794.5	21.7	681	2 Q9IT65	Q9IT65 pseudomonas
13	781.5	21.3	679	2 Q9CMX1	Q9CMX1 pasteurella
14	702	19.1	678	2 Q9K1E2	Q9K1E2 neisseria m
15	694	18.9	678	2 Q9JX57	Q9JX57 neisseria m
16	690.5	18.8	713	2 P74571	P74571 synecocyst
17	681.5	18.6	714	10 Q9LSL3	Q9LSL3 arabidopsis
18	671.5	18.3	685	2 Q9RTU7	Q9RTU7 deinococcus
19	635.5	17.3	723	10 Q9LXA6	Q9LXA6 arabidopsis

20	627.5	17.1	674	2 Q9PH19	Q9PH19 xylella fas
21	582	15.9	860	10 Q9FHN0	Q9FHN0 arabidopsis
22	578.5	15.8	716	2 Q9PC41	Q9PC41 xylella fas
23	556	15.2	699	5 Q9V9H7	Q9V9H7 drosophila
24	553	15.1	677	2 Q9A264	Q9A264 caulobacter
25	507.5	13.8	711	11 Q9CYN6	Q9CYN6 mus musculu
26	495.5	13.5	678	5 Q60982	Q60982 leishmania
27	478	13.0	93	11 Q9CUL3	Q9CUL3 mus musculu
28	452.5	12.3	730	5 Q9VC06	Q9VC06 drosophila
29	268.5	7.3	265	3 Q9P450	Q9P450 schizophyll
30	191.5	5.2	127	5 Q9NLT5	Q9NLT5 leishmania
31	189	5.2	820	5 Q62512	Q62512 caenorhabdi
32	175.5	4.8	615	2 Q9KQL5	Q9KQL5 vibrio chol
33	166	4.5	81	11 Q99LK5	Q99LK5 mus musculu
34	143	3.9	532	2 Q9KEV0	Q9KEV0 bacillus ha
35	141.5	3.9	2747	5 Q9BJX9	Q9BJX9 plasmodium
36	140	3.8	2166	2 Q51465	Q51465 borrelia bu
37	139.5	3.8	852	2 Q9X1X1	Q9X1X1 thermotoga
38	139.5	3.8	2269	5 Q62223	Q62223 plasmodium
39	139	3.8	1084	2 Q83423	Q83423 treponema p
40	137.5	3.7	880	1 Q9U2C8	Q9U2C8 pyrococcus
41	135	3.7	578	2 Q9ZM03	Q9ZM03 helicobacte
42	135	3.7	1095	5 Q61566	Q61566 ascaris suu
43	134	3.7	767	5 Q9V4F2	Q9V4F2 drosophila
44	134	3.7	3734	3 Q9C1G0	Q9C1G0 kalliichroma
45	133	3.6	547	2 Q9X058	Q9X058 thermotoga

ALIGNMENTS

RESULT 1  
Q9BYT8 PRELIMINARY; PRT; 704 AA.  
AC Q9BYT8  
DT 01-JUN-2001 (TREMREL. 17, Created)  
DT 01-JUN-2001 (TREMREL. 17, Last sequence update)  
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)  
DE NEUROLYSIN (EC 3.4.24.16).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Chen J.M., Rawlings N.D., Barrett A.J.;  
RT "Cloning and sequencing of human neurolysin, an oligopeptidase of family M3."  
RL Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL: AJ300837; CAC27329.1;  
KW Hydrolase.  
SQ SEQUENCE. 704 AA; 80651 MW; 80136688D79BBEDF CRC64;

Query Match	100.0%;	Score 3668;	DB 4;	Length 704;
Best Local Similarity	100.0%;	Pred. No. 2.4e-229;		
Matches 704;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1	MIARCLLAVSLRRVGGSKILLRMTLGRVMSPLQAMSSYTVAGRNVLRLWDLSPQIKTR	60		
Db 1	MIARCLLAVSLRRVGGSKILLRMTLGRVMSPLQAMSSYTVAGRNVLRLWDLSPQIKTR	60		
QY 61	TTELVQTVQVYDAVGMGLGIEEVTYENCLOALADVEKVIIVERTMLDFFQHVSSDEKRA	120		
Db 61	TTELVQTVQVYDAVGMGLGIEEVTYENCLOALADVEKVIIVERTMLDFFQHVSSDEKRA	120		
QY 121	ASTEADKRLSRFDIEMSRMGDIFERIVHLQETCDLGKIKPEARRYLEKSKMKGRNGLHL	180		
Db 121	ASTEADKRLSRFDIEMSRMGDIFERIVHLQETCDLGKIKPEARRYLEKSKMKGRNGLHL	180		
QY 181	PEQVQNEIKSMKRMSELCDNFKNLNEDDTFLVFSKAEALGALPDFIDSLKTDKDYK	240		
Db 181	PEQVQNEIKSMKRMSELCDNFKNLNEDDTFLVFSKAEALGALPDFIDSLKTDKDYK	240		

```
QY 241 ITLKYPHYFPMKKCCIPETRRRMEMAFNTRCKBENTIILOQLPLRTKVKAKLGYSTHA 300
DB 241 ITLKYPHYFPMKKCCIPETRRRMEMAFNTRCKBENTIILOQLPLRTKVKAKLGYSTHA 300
QY 301 DFVLEMTAKSTSRVTAFLDLSOKLPLGEAREFIIINLKKKECKDRGEYDGKINAWD 360
DB 301 DFVLEMTAKSTSRVTAFLDLSOKLPLGEAREFIIINLKKKECKDRGEYDGKINAWD 360
QY 361 LYYTWTQTEELKYSIDQEFLEKPEPIEVVTEGLLNTYQELLGLSFEQMTDAHVNKSVTL 420
DB 361 LYYTWTQTEELKYSIDQEFLEKPEPIEVVTEGLLNTYQELLGLSFEQMTDAHVNKSVTL 420
QY 421 YTVKDKATGEVLGYLDLXPREGKYHNAACFGLQPGCLLPDGRMMAVAALVYVNSQPV 480
DB 421 YTVKDKATGEVLGYLDLXPREGKYHNAACFGLQPGCLLPDGRMMAVAALVYVNSQPV 480
QY 481 AGPSLLRHDEVRTYFHFEGHVMHQICAOQDFARFSGTNNVETDFVEVPSQMLENNWWDVD 540
DB 481 AGPSLLRHDEVRTYFHFEGHVMHQICAOQDFARFSGTNNVETDFVEVPSQMLENNWWDVD 540
QY 541 SLRSLSKHYKDGSPADLLEKLVASRLVNTGLTLRQIVLSKYVDQSILHTNTSLDAASEY 600
DB 541 SLRSLSKHYKDGSPADLLEKLVASRLVNTGLTLRQIVLSKYVDQSILHTNTSLDAASEY 600
QY 601 AKYCEILGVAATPGTNNPATFGLAGGYDQYGYGLWSEVFSMDMFYSCFKKEGINNPE 660
DB 601 AKYCEILGVAATPGTNNPATFGLAGGYDQYGYGLWSEVFSMDMFYSCFKKEGINNPE 660
QY 661 VGMKYRNILKPGSLDGMMLHNFRLKREPNOKAFMLSRGLHAP 704
DB 661 VGMKYRNILKPGSLDGMMLHNFRLKREPNOKAFMLSRGLHAP 704

RESULT 2
P79433 PRELIMINARY; PRT; 745 AA.
AC P79433;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ENDOPEPTIDASE 24.16 (EC 3.4.-.-).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
ON NCBI_TaxID=9823;
RX [1]
RN SEQUENCE FROM N.A.
RP TISSUE=LIVER;
RC MEDLINE=97326108; PubMed=9182559;
RA Kato A., Sugitara N., Saruta Y., Hosoiri T., Yasue H., Hirose S.;
RT "Targeting of endopeptidase 24.16 to different subcellular
RT compartments by alternative promoter usage.";
RL J. Biol. Chem. 272:15313-15322(1997).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);
CC ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.
DR EMBL; AB000172; BAA19064.1; -
DR EMBL; AB000173; BAA19065.1; -
DR EMBL; AB000174; BAA19066.1; -
DR EMBL; AB000175; BAA19067.1; -
DR EMBL; AB000425; BAA19105.1; -
DR EMBL; AB000411; BAA19105.1; JOINED.
DR EMBL; AB000414; BAA19105.1; JOINED.
DR EMBL; AB000415; BAA19105.1; JOINED.
DR EMBL; AB000416; BAA19105.1; JOINED.
DR EMBL; AB000417; BAA19105.1; JOINED.
DR EMBL; AB000418; BAA19105.1; JOINED.
DR EMBL; AB000419; BAA19105.1; JOINED.
DR EMBL; AB000420; BAA19105.1; JOINED.
DR EMBL; AB000421; BAA19105.1; JOINED.
DR EMBL; AB000422; BAA19105.1; JOINED.
DR EMBL; AB000423; BAA19105.1; JOINED.
DR EMBL; AB000424; BAA19105.1; JOINED.
```

```
DR EMBL; AB000170; BAA19061.1; -
DR EMBL; AB000171; BAA19062.1; -
DR EMBL; AB000172; BAA19063.1; -
DR EMBL; AB000425; BAA19106.1; -
DR EMBL; AB000414; BAA19106.1; JOINED.
DR EMBL; AB000415; BAA19106.1; JOINED.
DR EMBL; AB000416; BAA19106.1; JOINED.
DR EMBL; AB000417; BAA19106.1; JOINED.
DR EMBL; AB000418; BAA19106.1; JOINED.
DR EMBL; AB000419; BAA19106.1; JOINED.
DR EMBL; AB000420; BAA19106.1; JOINED.
DR EMBL; AB000421; BAA19106.1; JOINED.
DR EMBL; AB000422; BAA19106.1; JOINED.
DR EMBL; AB000423; BAA19106.1; JOINED.
DR EMBL; AB000424; BAA19106.1; JOINED.
DR InterPro: IPR001130; Zn_MTpeptidse.
DR InterPro: IPR001567; Peptidase_M3.
DR Pfam: PF01432; Peptidase_M3; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR Metalloprotease; Hydrolase; Zinc.
SQ SEQUENCE 745 AA; 85109 MW; EC451B8EAD4AC28F CRC64;

Query Match 93.88; Score 3441; DB 6; Length 745;
Best Local Similarity 95.18; Pred. No. 1.3e-214;
Matches 656; Conservative 15; Mismatches 19; Indels 0; Gaps 0;
QY 15 VGGSRILLRTWLTGREVMSPLQAMSSYTVAGRNVLRWDLSPQIKTRTEELIVQTKQYDA 74
DB 56 VGGSGILLRMTLGREAMSPLOAMSSYTVAGNVLRWDLSPQIKRRTEELIAQTKQYDD 115
QY 75 VGMIGIEVITYENCLOALADVEVKYIVERTMLDPQHVSSDKVEVRAAESTEADKRLSPDI 134
DB 116 IGMLDIEVITYENCLOALADVEVKYIVERTMLDPQHVSSDKVEVRAAESTEADKRLSPDI 175
QY 135 EMSMRGDIPIRIVHLOETCDLGKIKPEARRYLEKSIKMGKENGHLDPQVQNEIKSMKKR 194
DB 176 EMSMRDIFLRIVRLKETCDLGKIKPEARRYLEKSVKMGKRNGLHLPQVQNEIKAMKKR 235
QY 195 MSELICIDFNKLNEDDTFLVFSKALGALPDFIDSLSEKTDDEDDKYKITLYKYPHYFPVKK 254
DB 236 MSELICIDFNKLNEDDTFLVFSKALGALPDFIDSLSEKTDDEDDKYKITLYKYPHYFPVKK 295
QY 255 CCIPETRRRMEMAFNTRCKBENTIILOQLPLRTKVKAKLGYSTHADFLVLEMTAKSTSR 314
DB 296 CCIPETRRRMEMAFNTRCKBENTIILOQLPLRAKVAKLLGYSTHADFLVLEMTAKSTHH 355
QY 315 VTAFLDLSOKLPLGEAREFIIINLKKKECKDRGEYDGKINAWDLYYTWTQTEELKYS 374
DB 356 VTAFLDLSOKLPLGEAREFIIINLKKKECKDRGEYDGKINAWDLYYTWTQTEELKYS 415
QY 375 IDQEFLEKPEPIEVVTEGLLNTYQELLGLSFEQMTDAHVNKSVTLTYVKDKATGEVLGQ 434
DB 416 VQDQELKPEPIEVVTEGLLNTYQELLGLSFEQMTDAHVNKSVTLTYVKDKATGEVLGQ 475
QY 435 FYLDLYPREGKYHNAACFGLQPGCLLPDGRMMAVAALVYVNSQPVAGRPSLLRHDEVRT 494
DB 476 FYLDLYPREGKYHNAACFGLQPGCLLPDGRMMSVAALVYVNSQPVAGRPSLLRHDEVRT 535
QY 495 YFHFEGHVMHQICAOQDFARFSGTNNVETDFVEVPSQMLENNWWDVDSLRLSKHYKDGSP 554
DB 536 YFHFEGHVMHQICAOQDFARFSGTNNVETDFVEVPSQMLENNWWDVDSLRLSKHYKDGSP 595
QY 555 IADDLLEKLVASRLVNTGLTLRQIVLSKYVDQSILHTNTSLDAASEYAKYCEILGVAATP 614
DB 596 ITDDLLEKLVASRLVNTGLTLRQIVLSKYVDQSILHTNTSLDAASEYAKYCEILGVAATP 655
QY 615 GTNNPATFGLAGGYDQYGYGLWSEVFSMDMFYSCFKKEGINNPEVGMKYRNILKPGG 674
DB 656 GTNNPATFGLAGGYDQYGYGLWSEVFSMDMFYSCFKKEGINNPEVGMKYRNILKPGG 715
QY 675 SLDGMMLHNFRLKREPNOKAFMLSRGLHAP 704
DB 715 SLDGMMLHNFRLKREPNOKAFMLSRGLHAP 704
```

Ds 716 SLDGMDLQNLKREPNOKAFLMSRGLHAP 745

```
RESULT 3
Q9BQD0 PRELIMINARY; PRT; 607 AA.
AC Q9BQD0;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE UNKNOWN (PROTEIN FOR MGC:3978) (SIMILAR TO THIMET OLIGOPEPTIDASE 1).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY ADENOCARCINOMA;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY ADENOCARCINOMA;
RA Strausberg R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004985; AAH04985.1;
DR EMBL; BC001644; AAH01644.1;
SQ SEQUENCE 607 AA; 70304 MW; 1CB7DEAF29A5B2C3 CRC64;

Query Match 72.0%; Score 2640; DB 4; Length 607;
Best Local Similarity 94.8%; Pred. No. 7.1e-163;
Matches 514; Conservative 4; Mismatches 14; Indels 10; Gaps 2;

QY 1 MTARCLAVRSRRVGGSKILLRMTLGRVMSPLQAMSYTVAGRNVLKRWLSPQIKTR 60
Db 1 MIARCLAVRSRRVGGSKILLRMTLGRVMSPLQAMSYTVAGRNVLKRWLSPQIKTR 60
QY 61 TEELIVQKQVDAVGMGLIEVTVYENCLQALADVEKVIYVERTMLDPQHVSSDKEYRA 120
Db 61 TEELIVQKQVDAVGMGLIEVTVYENCLQALADVEKVIYVERTMLDPQHVSSDKEYRA 120
QY 121 ASTEADKRLSRPDIEMSRGDIFFERIVHLQETCDLGKIKPEARRYLEKSIMKGRNGLHL 180
Db 121 ASTEADKRLSRPDIEMSRGDIFFERIVHLQETCDLGKIKPEARRYLEKSIMKGRNGLHL 180
QY 181 PQVONEIKSMKRMSELICIDPNKNLNEDDTPLVFSKAEGLALPDFIDSLKTDKDYK 240
Db 181 PQVONEIKSMKRMSELICIDPNKNLNEDDTPLVFSKAEGLALPDFIDSLKTDKDYK 240
QY 241 ITLKYPHYPPVMKKCCIPETRRMEMAFNTRCKEENTIILOQLPLRTKVAKLLGYSTHA 300
Db 241 ITLKYPHYPPVMKKCCIPETRRMEMAFNTRCKEENTIILOQLPLRTKVAKLLGYSTHA 300
QY 301 DFVLEMTAKSTSRVTAFLDLDLSQKLPLGEAEREFILNKKKECKDRGFYDVGKINAWD 360
Db 301 DFVLEMTAKSTSRVTAFLDLDLSQKLPLGEAEREFILNKKKECKDRGFYDVGKINAWD 360
QY 361 LYYMTQTEELKYSIDQBELFYFFIEVTVTEGLLNTYQELGLSFEQMTDAHVWVNSVT 420
Db 361 LYYMTQTEELKYSIDQBELFYFFIEVTVTEGLLNTYQELGLSFEQMTDAHVWVNSVT 420
QY 421 YTVKDKATGEVLGOYLDLYPREGYNHAACFGLPGCLLPDGRMMAVAALVVPSPQV 480
Db 421 YTVKDKATGEVLGOYLDLYPREGYNHAACFGLPGCLLPDGRMMAVAALVVPSPQV 480
QY 481 AGRPSLLRHDEVRTYFHEFGVHMQICATDFARFSGTNVETDF-----VEVPSQM 531
Db 481 AGRPSLLRHDEVRTYFHEFGVHMQICATDFARFSGTNVETDF-----VEVPSQM 531
QY 532 LE 533
Db 540 FQ 541
```

```
RESULT 4
Q9ULJ4 PRELIMINARY; PRT; 447 AA.
AC Q9ULJ4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE KIAA1226 PROTEIN (FRAGMENT).
GN KIAA1226.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
DR EMBL; AB033052; BAA86540.1;
DR InterPro; IPR000130; ZN_MTPeptidse.
DR InterPro; IPR001567; Peptidase_M3.
DR Pfam; PF01432; Peptidase_M3; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 447 AA; 50976 MW; 5A1E6DD76C00E040 CRC64;

Query Match 64.2%; Score 2355; DB 4; Length 447;
Best Local Similarity 100.0%; Pred. No. 1.3e-144;
Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 258 PSTRRMEAFNTRCKEENTIILOQLPLRTKVAKLLGYSTHADFLVLENTAKTSRVRTA 317
Db 1 PSTRRMEAFNTRCKEENTIILOQLPLRTKVAKLLGYSTHADFLVLENTAKTSRVRTA 60
QY 318 FLDDLSQLKPLGEAEREFILNKKKECKDRGFYDVGKINAWDLYYMTQTEELKYSIDQ 377
Db 61 FLDDLSQLKPLGEAEREFILNKKKECKDRGFYDVGKINAWDLYYMTQTEELKYSIDQ 120
QY 378 EPLKYPFIEVTVTEGLLNTYQELGLSFEQMTDAHVWVNSVTLYTVKDKATGEVLGOY 437
Db 121 EPLKYPFIEVTVTEGLLNTYQELGLSFEQMTDAHVWVNSVTLYTVKDKATGEVLGOY 180
QY 438 DLYPREGYNHAACFGLPGCLLPDGRMMAVAALVVPSPQVAGRPSSLRHDEVRTYF 497
Db 181 DLYPREGYNHAACFGLPGCLLPDGRMMAVAALVVPSPQVAGRPSSLRHDEVRTYF 240
QY 498 EFGVHMQICATDFARFSGTNVETDFVEVPQMLNENWVDVSLRRLSKHYKDGSPAD 557
Db 241 EFGVHMQICATDFARFSGTNVETDFVEVPQMLNENWVDVSLRRLSKHYKDGSPAD 300
QY 558 DLLEKLVASRLVNTGLLRQIVLSKVDSQSLTNTSLDAASEYAKYCSILGVAATPGTN 617
Db 301 DLLEKLVASRLVNTGLLRQIVLSKVDSQSLTNTSLDAASEYAKYCSILGVAATPGTN 360
QY 618 MPATGHLAGGYDGOYGYLVSEVFSMDMFYSCFKKEGIMNPEVGMKYNRLILKPGGSLD 677
Db 361 MPATGHLAGGYDGOYGYLVSEVFSMDMFYSCFKKEGIMNPEVGMKYNRLILKPGGSLD 420
QY 678 GMDMLHNLKREPNOKAFMLSRGLHAP 704
Db 421 GMDMLHNLKREPNOKAFMLSRGLHAP 447

RESULT 5
Q9EPX1 PRELIMINARY; PRT; 687 AA.
ID Q9EPX1
```

AC Q9EPX1; 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE THIMET OLIGOPEPTIDASE (EC 3.4.24.15).  
GN THOPI.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1].  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20538378; PubMed=10969067;  
RA Tullai J.W., Cummins P.M., Pabon A., Roberts J.L., Lopingco M.C.,  
RA Shrimpton C.N., Smith A.I., Martignetti J.A., Ferro E.S.,  
RA Gluckman M.J.;  
RT "The Neuropeptide Processing Enzyme EC 3.4.24.15 Is Modulated by  
RT Protein Kinase A Phosphorylation.";  
RL J. Biol. Chem. 275:36514-36522(2000).  
DR EMBL; AF314187; AAC35061.1; -;  
DR MGD; MG1:1354165; Thopi.  
DR InterPro: IPR001567; Peptidase.M3.  
DR InterPro: IPR000130; Zn\_MTPeptidse.  
DR Pfam: Pf01432; Peptidase.M3; 1.  
DR PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
KW Hydrolase.  
SQ SEQUENCE 687 AA; 77994 MW; F34A63D7C2A6D018 CRC64;

Query Match 62.7%; Score 2299; DB 11; Length 687;  
Best Local Similarity 64.8%; Pred. No. 1e-140;  
Matches 425; Conservative 92; Mismatches 139; Indels 0; Gaps 0;

Qy 46 NVLRWDLSPQIKTRTEELIVQTKQVDAVGMGLGIEEVTYENCLOALADVEVKIVERTM 105  
Db 22 NVLRWDLSPQIKTRTEELIVQTKQVDAVGMGLGIEEVTYENCLOALADVEVKIVERTM 81  
Qy 106 LDFQHVSSDEKRAASTADKRLSRDIEMSRGDIPIERIVHLEQTCDLGKIKPEARRY 165  
Db 82 LDFQHVSSDEKRAASTADKRLSRDIEMSRGDIPIERIVHLEQTCDLGKIKPEARRY 141  
Qy 166 LEKSIKMGKRNGLHLPQVQVNEIKSMKRMSELICIDFNKNLNEDDTFLVFSKAEGLALPD 225  
Db 142 LERLIKLRNGLHLPQVQVNEIKSMKRMSELICIDFNKNLNEDDTFLVFSKAEGLALPD 201  
Qy 226 DFIDSLEKTDGKYNLKYTPHYFPVMKKCCIPETRRMEMAFNTRCKEENTIILOQLLP 285  
Db 202 DFIDSLEKTDGKYNLKYTPHYFPVMKKCCIPETRRMEMAFNTRCKEENTIILOQLLP 261  
Qy 286 LRTKVALIGYSTHADPVLNMTAKTSRVTAFDLDLSQKLPLGEAREFERILNKKKEC 345  
Db 262 LRAQKSSILGPHTHADPVLNMTAKTSRVTAFDLDLSQKLPLGEAREFERILNKKKEC 321  
Qy 346 KDRGFEDGKINAWDLYYMTQTEELKYSIDQEFLEKYEFPVEVTEGLLNTYQELLGLSF 405  
Db 322 AKRGLPFDGRHAWDMRYMNMQVEETRYRVDQNLKKEFPVQVYTRGLATYQELLGLTF 381  
Qy 406 EQMTDAHWNKSVTLTYVKDKATGEVLGQFYLDLYPREGKYNHAAACFGLOPGCLLPDGSR 465  
Db 382 TLEGAAAHWEDVRLYSVRDASGEEIGKFFVLDLYPREGKYNHAAACFGLOPGCLLPDGSR 441  
Qy 466 MMVAALVNFNSQVACRPSLLRHDEVRTYFHEFGVHMHQICAOITDFARFSGTINVTDFV 525  
Db 442 OLAVAAMVNFNSQVACRPSLLRHDEVRTYFHEFGVHMHQICAOITDFARFSGTINVTDFV 501  
Qy 526 EVPSOMLENWVDVSLRLSKYKDGSPADDDLEKLVASRLVNTGLLTIRQIVLSKVD 585  
Db 502 EAPSOMLNWWERELMRMSQHYRTGGEAPQDLLEKLVASRLVNTGLLTIRQIVLSKVD 561  
Qy 586 QSLHTNTSLDAASAYKVCSEILGVAATPGNNPATGHLAGGAGGYGYGLWSEVFSMD 645  
Db 562 QVLTHTQTDADPAEYARLCQELGVPATPGNNPATGHLAGGAGGYGYGLWSEVFSMD 621

Qy 646 MFYSCFPKKEGIMNPVGMKYRNILILKPGSLDGMDFLHFLKREPNOKAFILMSRGL 701  
Db 622 MFHTRFKQEGVLSPKVGMDYRTSILRPGSSDASAMILKQFLGRDPKQDAFLLSKGL 677

RESULT 6  
Q9PTV2 PRELIMINARY; PRT; 685 AA.  
AC Q9PTV2;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE THIMET OLIGOPEPTIDASE.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1].  
RP SEQUENCE FROM N.A.  
RC TISSUE=OVARY;  
RA Okida N., Ishikawa K., Tokumoto M., Tokumoto T.;  
RT "Xenopus thimet oligopeptidase.";  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB030904; BAA89309.1; -;  
DR InterPro: IPR00130; Zn\_MTPeptidse.  
DR InterPro: IPR001567; Peptidase.M3.  
DR Pfam: Pf01432; Peptidase.M3; 1.  
DR PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
SQ SEQUENCE 685 AA; 78137 MW; F41C5EE45BBB78ED CRC64;

Query Match 62.5%; Score 2291.5; DB 13; Length 685;  
Best Local Similarity 64.4%; Pred. No. 3.1e-140;  
Matches 426; Conservative 98; Mismatches 135; Indels 3; Gaps 1;

Qy 46 NVLRWDLSPQIKTRTEELIVQTKQVDAVGMGLGIEEVTYENCLOALADVEVKIVERTM 105  
Db 19 NVLRWDLTPKQIEITLGEVTEQTKRVYDQVGSLLSSVTYDNTLKLVLADVEVSGKRS 78  
Qy 106 LDFQHVSSDEKRAASTADKRLSRDIEMSRGDIPIERIVHLEQTCDLGKIKPEARRY 165  
Db 79 LDFQHVSSDEKRAASTADKRLSRDIEMSRGDIPIERIVHLEQTCDLGKIKPEARRY 138  
Qy 166 LEKSIKMGKRNGLHLPQVQVNEIKSMKRMSELICIDFNKNLNEDDTFLVFSKAEGLALPD 225  
Db 139 LERVIKLSQRNGLNLPSTQEKIKSIKKKSTLSDFNKNLNEDDTFLVFSKAEGLALPD 198  
Qy 226 DFIDSLEKTDGKYNLKYTPHYFPVMKKCCIPETRRMEMAFNTRCKEENTIILOQLLP 285  
Db 199 DFLNSLEKAGDGKILKILKYTPHYFPVMKKCCIPETRRMEMAFNTRCKEENTIILOQLLP 258  
Qy 286 LRTKVALIGYSTHADPVLNMTAKTSRVTAFDLDLSQKLPLGEAREFERILNKKKEC 345  
Db 259 LREGNSSILGPHTHADPVLNMTAKTSRVTAFDLDLSQKLPLGEAREFERILNKKKEC 318  
Qy 346 KDRGFEDGKINAWDLYYMTQTEELKYSIDQEFLEKYEFPVEVTEGLLNTYQELLGLSF 405  
Db 319 KKRGLNFDQINAWDMRYMNMQVEETRYRVDQNLKKEFPVQVYTRGLATYQELLGLNF 378  
Qy 406 EQMTDAHWNKSVTLTYVKDKATGEVLGQFYLDLYPREGKYNHAAACFGLOPGCLLPDGSR 465  
Db 379 VLEKGAAMHEDVRLYSVRDATTGQIKGYLDLYPREGKYNHAAACFGLOPGCLLPDGSR 438  
Qy 466 MMVAALVNFNSQVACRPSLLRHDEVRTYFHEFGVHMHQICAOITDFARFSGTINVTDFV 525  
Db 439 QISVAAMVNFNSQVACRPSLLRHDEVRTYFHEFGVHMHQICAOITDFARFSGTINVTDFV 498  
Qy 526 EVPSOMLENWVDVSLRLSKYKDGSPADDDLEKLVASRLVNTGLLTIRQIVLSKVD 585  
Db 499 EAPSOMLNWWERELMRMSQHYRTGGEAPQDLLEKLVASRLVNTGLLTIRQIVLSKVD 558  
Qy 586 QSLHTNTSLDAASAYKVCSEILGVAATPGNNPATGHLAGGAGGYGYGLWSEVFSMD 645

```

Db 559 QVLTQIGVDPVEEYKSLNSNETLIGIPATPGTNMLSAFAHLAGGYDAQYGYLWSEVYSND 618
Qy 646 MYSCFKKEGINPVGKYNRLILKPGSLDGMDLNHLNFKRPNQKAFILMSRGL---H 702
Db 619 MYTFKQEGIMSKKVGADYRNCILKPGSLDGMDLNHLNFKRPNQKAFILMSRGLANVN 678
Qy 703 AP 704
Db 679 AP 680

RESULT 7
ID Q9BW75 PRELIMINARY; PRT; 480 AA.
AC Q9BW75;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE TIMET OLIGOPEPTIDASE 1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=NEUROBLASTOMA;
RA Strausberg R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC000583; AA000583.1; -.
SQ SEQUENCE 480 AA; 54735 MW; 96250597AC877913 CRC64;
```

```

Query Match 46.4%; Score 1703; DB 4; Length 480;
Best Local Similarity 65.5%; Pred. No. 2.4e-102;
Matches 306; Conservative 68; Mismatches 93; Indels 0; Gaps 0;

Qy 235 DDYKYLTKLPHYPVPMKCCIPETRRMEMAFNTRCKEENIILQQLPLRTKVKALL 294
Db 2 EDGKLKVLTKLPHYPVPMKCCIPETRRMEMAFNTRCKEENIILQQLPLRTKVKALL 294
Qy 295 GYSTHADFVLEMTAKTSRVTAFDDLSQKLPLGEAREFILLKKECKDRGFEDG 354
Db 62 GPHTHADVYLEMNAKTSOTVATFDELAKQLPLGEAREFILLKKECKDRGFEDG 121
Qy 355 KINANDLYYMTQTEELKYSIDQELKYEFPVETVTEGLLNTYQELGLSPQMTDAHV 414
Db 122 RIRAWDMRYNNQVETRYCDONLLKEYFPVQVYVTHGLLGYQELGLAFHHEGASAW 181
Qy 415 NKSVTLYTKDKATGEVLGQFVLDLYPREGKYNHAACFGLQPGCLLPDGRMMVAALV 474
Db 182 HDVRLYIARDASGEVVGKFLVDLYPREGKYGHACFGLQPGCLLPDGRMMVAALV 241
Qy 475 NFSQVAGRPSLLRHDEVRTYFHEFGHVMHQICATDFARFSGTNVETDFVEVPQMLEN 534
Db 242 NFKTPTADAPSLQLQDEVTYFHEFGHVMHQICATDFARFSGTNVETDFVEVPQMLEN 301
Qy 535 WYWDVDSLRRLSKHYKDGSPITADLLEKLVASRLVNTGLLTLRQIVLSKVDSLSHTSL 594
Db 302 WYWEQEPRLRSRHYRTGSAPVRELEKLIERSQANTGLNLRQIVLAKVQDQALHTQDA 361
Qy 595 DAASVYAKVCSILGVAATPGTNMPATFGLAGGYDGYGYGLWSEVSDMFYSCFKKE 654
Db 362 DPAEYARLCQELQIVPATPGTNMPATFGLAGGYDGYGYGLWSEVSDMFYSCFKKE 421
Qy 655 GIMNPVGKYNRLILKPGSLDGMDLNHLNFKRPNQKAFILMSRGL 701
Db 422 GVLNSKVGMDYRSCILRPGGSDASAMLRRLRFLGRDPKQDAFLLSKGL 468

RESULT 8
ID P97996 PRELIMINARY; PRT; 716 AA.
Qy P97996
```

```

AC P97996;
DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE INTRACELLULAR METALLOPROTEINASE MEPB.
GN MEPB.
OS Aspergillus fumigatus (Sartorya fumigata).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5085;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97388579; PubMed=9245813;
RA Ibrahim-Granet O., d'Enfert C.;
RT "The Aspergillus fumigatus mepb gene encodes an 82 kDa intracellular
RT metalloproteinase structurally related to mammalian thimet
RT oligopeptidases."
RL Microbiology 143:0-0(0).
DR EMBL; U85769; AAB66656.1; -.
DR MEROPS; M03.009; -.
DR InterPro; IPR001567; Peptidase_M3.
DR InterPro; IPR000130; Zn_Mrpeptidse.
DR Pfam; PF01432; Peptidase_M3; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 716 AA; 81923 MW; 6480A86C90185322 CRC64;
```

```

Query Match 32.1%; Score 1177; DB 3; Length 716;
Best Local Similarity 37.7%; Pred. No. 4.9e-68;
Matches 273; Conservative 125; Mismatches 253; Indels 74; Gaps 15;

Qy 33 PLQAMSSYTVAGRNVLRLWDLSPQIKTRTEELIVOTKVQVD-AVGMLGTEETVENCLOA 91
Db 9 PPQAPPLTATAQSI-----DDAKRLIETSRKVRDINIVANKPESATFDSVLAP 58
Qy 92 LADVEKVIIVERTMLDFPHQVSDKEVRAASTADKRLSRFDIEMSRGDI-----ERIVH 148
Db 59 LAHDENTMALESHILSFYQAVSTEQELRDASSKABELMDEFFIETVMREDVFKLDAVLK 118
Qy 149 LOETCDLGIKPEARYLEKSIKMKRNLGLHPE-QVQNEIKSMKMRKMSLCLIDFNKLN 207
Db 119 KNEISL-----PESRLLEKEHKYIRNGLGLPAGPKDRFKRKEIKRLSQISIEFOKNL 173
Qy 208 EDDTLVFSKAEALGALPDDFIDSLKTD--DDKYKITLKYPHYPPVPMKKCCIPETRRM 264
Db 174 ENGGINFTRELDGVPEDVLSGLKKEGEGEKWLFPKYDLPFTMYAKNAETRRKL 233
Qy 265 ENAFNTRKEENTIILOQLPLRTKVKALLGYSTHADFVLEMTAKTSRVTAFDLDLSQ 324
Db 234 MIENENKC-NQNVPLFREAVILRDEAARLLGYPNHAAFRIEDKMAKTPKTVDDFLGLDLS 292
Qy 325 KLKPLGEAREFILLKKECKDRGFEDGKINANDLYYMTQTEELKYSIDQELKYEFP 384
Db 293 RLTAGGHKEIKALLEKRADESERGEFPDGRYLYLDHREYDLMLLEKDYSLDQQLIAEYF 352
Qy 385 PLIEVTEGLLNTYQELGLSPQMTDAH-----VNNKSVTLTYV-KDKATGE 430
Db 353 PLQTTIEGMLKIFEEFLGLVFEITGEDREKVAPTKGSDIVWHEDVQVFSVWMDGEGS 412
Qy 431 -VLGQFYLDLYPREGKYNHAAFCGLQPGCLLPDGRMMVAALVNFSPQVAGRPSLLRH 489
Db 413 GFVGYLILDLFFPRPKYGHAAFNLPQPGIDAGKRRYPATALVCNFTKPTAKKPSLLAH 472
Qy 490 DVRTYFHEFGHVMHQICATDFARFSGTNVETDFVEVPQSMLENWVDSLRRLSKHY 549
Db 473 DEWTLFELHGLGHIHDLVSRITYSRFTGHTSTVRDFEAPQSMLENWCVTPSOLSKSKHY 532
Qy 550 KDGSP-----IADDLLEKLVASRLVNTGLLTLRQIVLSKVDSLSHT 590
Db 533 STLSPEYLAGWQEQANGKPPQPPERIPDEIENLIRTKHVNDALFNLRQLHFGIFDMTHVE 592
Qy 591 NTS-----LDAASEVAKYCS-----EILGVAATPGTNMPATFGLHAGGYDGYGY 636
```

Db 593 AESHEQIQKLPSTYTNQLRKQIALQDGPVGLMGDEWG-HGEATFGHLGGYDAGYGY 651  
QY 637 LWSVEFMDMYSCFKKGINNPGVGMKYNRLILKPGSLDGMMLHNFILKREPNQAF 696  
Db 652 LSSQVYSDMYETVTKDDP-MNPAAGRRYRGVLEKSGSQDEMTKLTDFLGREPTDAFY 710  
QY 697 MSRGL 701  
Db 711 KEGL 715

RESULT 9  
Q9FXD5 PRELIMINARY; PRT; 998 AA.  
AC Q9FXD5;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE F12A21.16.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Chao Q., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,  
RA Shinn P., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,  
RA Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B., Lee J.,  
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,  
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,  
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,  
RA Ecker J.R.;  
RT "Genomic sequence for Arabidopsis thaliana BAC F12A21 from chromosome  
I.";  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Chao Q., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,  
RA Shinn P., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,  
RA Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B., Lee J.,  
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,  
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,  
RA Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N., Theologis A.,  
RA Ecker J.R.;  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,  
RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,  
RA Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N.,  
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,  
RA Theologis A., Ecker J.R.;  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP EMBL; AC081113; AAC28905.1; -  
DR InterPro: IPR001567; Peptidase\_M3.  
DR InterPro: IPR000130; Zn\_MTPetdse.  
DR Pfam: PF01432; Peptidase\_M3; 1.  
DR PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN\_1.

SQ SEQUENCE 998 AA; 112949 MW; 49CBE1C8225955E8 CRC64;  
Query Match 27.7%; Score 1017.5; DB 10; Length 998;  
Best Local Similarity 33.7%; Pred. No. 1.7e-57;  
Matches 221; Conservative 157; Mismatches 244; Indels 33; Gaps 9;  
QY 51 DLSPEQIKTRTEELIVQTVKDYAVGMGLGIBEVTYENCALQADAVEVKYIIVERTMLDPQ 110  
Db 40 NLSAKEILDIAEEIIHKSTRVHDVALVSLDKSYENVVPLAELEARQLSLIOCCVFPK 99  
QY 111 HVSSDKVEVRAASTEADKRLSRFDIEMSRMGDIFERIVHLQETCDLGKIKPARYLEKSI 170  
Db 100 MLSPHDNVRKASTEAEQKIDAHILSCRKREDVY-RIIKIY-AARGESISPAAKCYLQCLV 157  
QY 171 KMGKRGHLHPEQVQNETSKMKRWSLCLIDFNKNLNEDDDFLVESKALGALPDDFTDS 230  
Db 158 RFDENGNLNTAIKREVERUKYIEIDLSLRYIIONLEDSSCLFFTEDELGLPFLQJON 217  
QY 231 LEKTDGDKYKITLKYPHYFPMKKCCIPETRRRMEMAFNTRCKEENTIILOQLPLRTKV 290  
Db 218 LEKTONKEFKLTLESRHVAALILECKIAKTRKTVMAYGKRCGDTNPVQLRQVSRHRL 277  
QY 291 AKLLGYSTHADFVLEMTAKTSRTVAPLDLSOKLPLGEAREREFILNLKKCKECKDRGF 350  
Db 278 ACVGYAHFADYALDRRMSKTSMRVIRFLEDISSSLTDLAIRFESILEDLKRKE----- 331  
QY 351 EYDGKI--NAWDLYYYMTOTEEELKYSIDQELKEYFPEVTVTEGLLNTYQELLGLSPEOM 408  
Db 332 --EGEIPFGVEDLLYIKRVEELQFDLDGDIRQIFPVNLVLSGFIKICQDLFGIKFEV 389  
QY 409 TDAHVWKSVTLYTVKDKATGEVLGQFVLDLYPREGYNHAAACFGQLPGCLLPDGSRRMA 468  
Db 390 TEVDVWYHDIRAFVAFVDSGSGKLLGYFLDMFTREGKNCVVALQNNALFSNGACQIP 449  
QY 469 VAALVNFSPQVAPRPSLLRHDEVRTTFEFGHVHMOICATDFARSGTNVETDFVEVP 528  
Db 450 VALLIAQFAKDGSGEAVPLGFSVDVNLHFHFGVHVHICNRASFARFSGLRVDPDFEIP 509  
QY 529 SOMLENWMDVDSLRRLSKHYKDGSPADLLEKLVASRLVNTGLTLRQIVLSKVQDSL 588  
Db 510 SOLLENW-----QDITK-----PLVDEVCKYLKWRYSFSAKLSQELIYCLFDQII 556  
QY 589 HTNTSLD-----AASEYAKYCEILGVAATPGTNMPATFGHLAGGYDGOYGYLWSEVPSM 644  
Db 557 YSDDADLQLIRSLHPKV--MIGLPVVECTNPASCFPRRAVIGSEATCYSLRWSEVAA 613  
QY 645 DMFVSCFKKEGIMNPEVGMKYNRLILKPGSLDGMMLHNFILKREPNQAFELMSR 699  
Db 614 DIFASKF-GDGHPLNYAGLOFRDKVLAPGGGKPEMLLTNFGREPSQAFIASR 667

RESULT 10  
Q9KVF8 PRELIMINARY; PRT; 680 AA.  
AC Q9KVF8;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE OLIGOPEPTIDASE A.  
GN VC0188.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=566;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-EL TOR N16961 / SEROTYPE O1;  
RX MEDLINE=20406833; PubMed=10952301;  
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,









Query Match		18.9%	Score 694;	DB 2;	Length 678;
Best Local Similarity		28.2%	Pred. No. 8.5e-37;		
Matches 190;		Conservative 131;	Mismatches 309;	Indels 44;	Gaps 18;
Qy	55	EQIKTRTEELIVQT--KQVYDAVGMGLGIEEV-YENCLOALADVEVYIIVERT-----ML	106		
Db	16	DOIKTEIDKPAQTAAIAEAREQIAAKAQHTGWANTVEPLTG-----ITERVGRINGVV	70		
Qy	107	DFPHVSDKEVRAASTADKRLSRDFDIEMSRGDIIFERIVHLOETCDLGKIRPEARLYL	166		
Db	71	SHLSNVTDTPELRAAYNELMPEITVFTEIGQDIELYNRFKTIKNSPFDTLSHAQTKL	130		
Qy	167	EKSIKMGKRNCLHPEQOVONEIKSMKKRMSLCIDFNKN-LNEDDTELVF--SKAELGAL	223		
Db	131	NHDLURDFVLSGAELPPQQAEALAKLTQEGAQLSAKFSQNVLDATDAFGIYFDDAPLAGI	190		
Qy	224	PDD----FIDSLEKTDDDKYKITLKYPHYFPVMKKCCIPETRRMEMAFNTRCKE-----	274		
Db	191	PEDALAMFAAAQSEGKTYKIGIQIPHYLAVIQYADNRKLRQIYRAYVTRASELSDDG	250		
Qy	275	--ENTIILOQLLPRTKVKALLGYSTHADVFLEMTAKSTSRVTAFLDDLSQKLPLGEA	332		
Db	251	KFDNTANIDRTLENALQTKLLGPKNYAELSLATKMADTPEQVLNLFHDLARRAKPYAEK	310		
Qy	333	EREFILNKKKECKDRGFEYDGKINADLYYYMTQTTELKYSIDQEFLEKXYFPTEVVTEG	392		
Db	311	D--LAEVKAFARESLGL---ADLPWDLGYAGEKLEAKYAFSETEVKKYFPVKVING	364		
Qy	393	LNTYQELLGLSFEQMTDAHWNKSVTLTYTVKDKATGEVLGQFYLDLYPREGKYNHAACF	452		
Db	365	LFAQIKKLYGIGFTEKT-VPVWHKDVRYFELQ--NGETIGGVYMDLYAREBKRGGAWN	421		
Qy	453	GLQPGCLLPDGSRRMAVAALVYNFSQPVAGRPSSLRLRHDEVRTYEHFEGHVMHQICAQTF	512		
Db	422	DYKGRRRFSDGTQLPTAYLVCNFTPPVGGKEARLSHDEILTLFHTGCHGLHLLTQVDE	481		
Qy	513	ARFSGTN-VETDFEVPSONLENWVDVDSLRRLSKHYKDGSPIDADDLLEKLVASRLVNT	571		
Db	482	LGVSGINGVEWDVDELPSQFMENFWWEYNVLAQMSAHEETGVPLPKELFDKMLAAKNFOR	541		
Qy	572	GLLTIRQIVLSKVDQSLHTNTSLDAASEYAKYCSFI---LGVAATPGTNMPA-TFGHL-A	626		
Db	542	GMELVRQMEFALDMMIYSEDEGRLKNWQVLDVSRKEVAVVRPPEYNRRFANSFGHIFA	601		
Qy	627	GGYDQYGYGLWSEVFSMDMYSCFKKEGINPVEGMKYRNLIILKPGSLDGMMDLHNFL	686		
Db	602	GGYSAGYYSYAAAEVLSADA-YAAFEESDDV-AATGKRFWQEIILAVGGSRAAESFKAFR	659		
Qy	687	KREPQKAFMSRG 700			
Db	660	GREPSIDALLRHSG 673			

Search completed: January 14, 2002, 19:02:18  
Job time: 377 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2002, 19:02:21 ; Search time 198.95 seconds  
(without alignments)  
9114.063 Million cell updates/sec

Title: US-09-833-782-1

Perfect score: 2115

Sequence: 1 atgatgcgcggcgccctttt.....gagccctgcatgctygtga 2115

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	2109.4	99.7	2864	22 AAF84499	Human aminopeptidase
2	1590.8	75.2	2148	13 AAQ27307	Rabbit endopeptidase
3	1178.6	55.7	1658	22 AAH14379	Human cDNA sequenc
4	790.8	37.4	2070	13 AAQ24875	Human amyloidin pr
5	618.4	29.2	776	22 AAH06280	Human cDNA clone (
6	339	16.0	339	21 AAH31048	Human colon cancer
7	308.8	14.6	900	21 AAC98797	Human pancreatic c
8	117	5.5	1830121	17 AAT42083	Haemophilus influe
9	112.8	5.3	936	22 AAF58252	Oligonucleotide D1
10	112.8	5.3	936	22 AAF58254	Oligonucleotide D1
11	112.8	5.3	936	22 AAF58257	Oligonucleotide D1

12	112.8	5.3	936	22 AAF58259	Oligonucleotide D2
13	112.8	5.3	936	22 AAF58262	Oligonucleotide D2
14	112.8	5.3	938	22 AAF58255	Oligonucleotide D1
15	111.8	5.3	936	22 AAF58252	Oligonucleotide D1
16	111.8	5.3	936	22 AAF58254	Oligonucleotide D1
17	111.8	5.3	936	22 AAF58257	Oligonucleotide D1
18	111.8	5.3	936	22 AAF58259	Oligonucleotide D2
19	111.8	5.3	936	22 AAF58262	Oligonucleotide D2
20	111.8	5.3	938	22 AAF58255	Oligonucleotide D1
21	94.8	4.5	588	22 AAH11342	Human cDNA clone (
22	75	3.5	2037	21 AAF21572	N. meningitidis pa
23	75	3.5	2037	21 AAA81287	N. meningitidis pa
24	75	3.5	2037	21 AAZ53270	Neisseria meningit
25	75	3.5	2037	21 AAZ54604	Neisseria meningit
26	75	3.5	5253	21 AAA81478	N. meningitidis pa
27	75	3.5	349980	21 AAF21544	Neisseria meningit
28	75	3.5	1437668	21 AAA81490	N. meningitidis B
29	73.4	3.5	1035	21 AAZ53267	Neisseria meningit
30	73.4	3.5	1035	21 AAZ54601	Neisseria meningit
31	73.4	3.5	1431	21 AAF21569	N. meningitidis pa
32	73.4	3.5	1431	21 AAA81284	N. meningitidis pa
33	71.2	3.4	623	21 AAF07718	Fusarium venenatum
34	70.2	3.3	2037	21 AAF21571	N. meningitidis pa
35	70.2	3.3	2037	21 AAF21574	N. meningitidis pa
36	70.2	3.3	2037	21 AAA81286	N. meningitidis pa
37	70.2	3.3	2037	21 AAA81289	N. meningitidis pa
38	70.2	3.3	2037	21 AAZ53268	Neisseria meningit
39	70.2	3.3	2037	21 AAZ53271	Neisseria meningit
40	70.2	3.3	2037	21 AAZ54603	Neisseria meningit
41	70.2	3.3	2037	21 AAZ54606	Neisseria meningit
42	67.4	3.2	465	22 AAI14810	Probe #4743 for ge
43	67.4	3.2	465	22 AAI36168	Probe #4854 used t
44	67.4	3.2	465	22 AAI04602	Probe #4593 used t
45	66.4	3.1	2382	21 AAC50567	Arabidopsis thalia

#### ALIGNMENTS

RESULT 1

AAF84499  
ID AAF84499 standard: cDNA; 2864 BP.

XX  
AC AAF84499;

XX  
DT 25-JUL-2001 (first entry)

XX  
DE Human aminopeptidase 22196-encoding cDNA.

XX  
KW Human: aminopeptidase 22196; neutral zinc metalloproteinase;

KW metalloenzyme; analgesic; cytostatic; lung cancer; colon cancer; tumour;

KW pain; drug screening; spleen disorder; infectious disease;

KW immune disorder; lung disorder; colon disorder; liver disorder;

KW uterine disorder; brain disorder; skin disorder; cardiac disorder;

KW vascular disorder; bone disorder; gene therapy; ss.

XX  
OS Homo sapiens.

XX  
FH Key Location/Qualifiers

FT CDS 300..2414

FT /\*tag= a

FT /product= "Human aminopeptidase 22196"

XX  
PN WO200123590-A2.

XX  
XX

XX  
PD 05-APR-2001.

XX  
XX

XX  
PF 02-OCT-2000; 2000WO-US27214.

XX  
XX

XX  
PR 30-SEP-1999; 99US-0409180.

XX  
XX

XX  
PA (MILL-) MILLENNIUM PHARM INC.



```

QY 1501 caagtgcacacagatttgcacagactgattttgcagatttagcggaacaaatgtg 1560
Db 1800 cactgtagtcacacagatttgcacagactgattttgcagatttagcggaacaaatgtg 1859
QY 1561 gaaactgactttgtagagtgccatcgcaaatgcttgaaattgggtgtggagctcgat 1620
Db 1860 gaaactgactttgtagagtgccatcgcaaatgcttgaaattgggtgtggagctcgat 1919
QY 1621 tccctcgaagattgtcaaaacattataaagatggaagccctattgcagacgactgtctt 1680
Db 1920 tccctcgaagattgtcaaaacattataaagatggaagccctattgcagacgactgtctt 1979
QY 1681 gaaaaactgttgccttagctgtccaacacaggtcttctgaacctgcgccagattgtt 1740
Db 1980 gaaaaactgttgccttagctgtccaacacaggtcttctgaacctgcgccagattgtt 2039
QY 1741 ttgagcaaatgtgacagctctcttcataccaacacacatgctggtgacagtgaatat 1800
Db 2040 ttgagcaaatgtgacagctctcttcataccaacacacatgctggtgacagtgaatat 2099
QY 1801 gccaaatactgcagaaatattagagttgcagctactccaggcacaataatgccagct 1860
Db 2100 gccaaatactgcagaaatattagagttgcagctactccaggcacaataatgccagct 2159
QY 1861 acctttggacatttgcagggggatacagatggccaatatattgatatctttggagtga 1920
Db 2160 acctttggacatttgcagggggatacagatggccaatatattgatatctttggagtga 2219
QY 1921 gtattttccatggatatgttttacagctgttttaaaaaagagggaataatgaatccagag 1980
Db 2220 gtattttccatggatatgttttacagctgttttaaaaaagagggaataatgaatccagag 2279
QY 1981 gttggaatgaaatacagaacaaactaatcttgaaacctgggggatctcttgacggcaatggac 2040
Db 2280 gttggaatgaaatacagaacaaactaatcttgaaacctgggggatctcttgacggcaatggac 2339
QY 2041 atgctccacaatttctgaaacgtgagccaaacccaaagccttctaatgattgagagc 2100
Db 2340 atgctccacaatttctgaaacgtgagccaaacccaaagccttctaatgattgagagc 2399
QY 2101 ctgcatgctcyctga 2115
Db 2400 ctgcatgctccgtga 2414

RESULT 2
ID AAQ27307
AC AAQ27307 standard; cDNA; 2148 BP.
XX
AC AAQ27307;
XX
DT 03-FEB-1993 (first entry)
XX
DE Rabbit endopeptidase cDNA (preliminary sequence).
XX
KW microsome; metalloproteinase; polymerase chain reaction; ss.
XX
OS Oryctolagus cuniculus.
XX
FH Key Location/Qualifiers
CDS 11..2137
FT /*tag= a
FT /product= endopeptidase
FT /note= "the identity of many of the nucleotides
is undefined"
XX
PN W09213080-A.
XX
PD 06-AUG-1992.
XX
PF 28-JAN-1992; 92WO-US00731.
XX

```

```

PR 28-JAN-1991; 91US-0646997.
XX (UNIW ) UNIV WASHINGTON.
XX
PI Davie BW, Kawabata S;
XX
DR WPI: 1992-284665/34.
DR P-PSDB; AAR26114.
XX
PT Microsomal endo-peptidase isolated from rabbit liver - used to
PT cleave protein substrates in-vivo and in-vitro
XX
PS Example 3; Fig 5; 48pp; English.
XX
CC The rabbit endopeptidase was purified from rabbit liver and
CC digested with CNBr. Families of degenerate primers were designed
CC based on the amino acid sequences of the CNBr fragments. The
CC primers were used to amplify sequences from a rabbit cDNA library.
CC An amplified fragment was isolated and used to probe the same
CC library. Two overlapping clones produced a 2148bp sequence. A
CC preliminary sequence analysis was performed in which much of the
CC 5' and 3'-most sequences are unassigned; the actual amino and
CC carboxyl terminals of the protein have not yet been determined.
CC The short untranslated 5' and 3' sequences are primer sequences.
CC See AAR26107-R26113 and AAQ27305-Q27307.
XX
SQ Sequence 2148 BP; 568 A; 452 C; 551 G; 506 T; 71 other;

```

```

Query Match 75.2%; Score 1590.8; DB 13; Length 2148;
Best Local Similarity 87.2%; Pred. No. 0;
Matches 1728; Conservative 1; Mismatches 247; Indels 6; Gaps 1;

QY 1 atgategcccggtgctcttggctgtgcgaagcctccgcagagttggtgttccaggatt 60
Db 167 atnategcccggtgcttcttccgtgtncgaggcctccacagggtgtgtgttccaggatt 226
QY 61 ttactcagaatcacgttagaagaagaagtgtctctcttcagcagaatgtcttctctat 120
Db 227 ctattcaaaatgacatcaggagaagaagtgtgtcaccctcttcacgcagcttcttctat 286
QY 121 actgtggctggcagaaatgttttaagatggatctttccaccagagcaaatataaacaaga 180
Db 287 actgcagctggcaggaatgttttaagatggatctttccaccagagcaaatcaagaaga 346
QY 181 actgaggagctcattgtgcagaccacaaacgggtgaagctgtgttggaaatgctcgatt 240
Db 347 actgaggagctcattgtgcagaccacaaacgggtgaatgtattctgttggaaatgctggatc 406
QY 241 gaggaagttaacttacagagaactgtctgcagcactgtgcagatgtagaagtaagtata 300
Db 407 aaggacgtgacttacagagaactgtctgcaggcgcgtgcagatgtggaagtgaagtacata 466
QY 301 gtggaaaggagaccatgctagactttcccccagcatgtatcctctgacaaagaagtacagca 360
Db 467 gtggaaagaaccatgctagactttcccccagcatgtttctactgacagagaagtacaggca 526
QY 361 gcaagtacagaagcagacaaaaagactttctcgtttttgattgattgagatgagatgagaga 420
Db 527 gcaagtacagaagcagcacaagaggctttctcgtttttgattgattgagatgagatgagaga 586
QY 421 gatattttgagagaattgttcatattacaggaacactgtgactctggggaagataaacct 480
Db 587 gatattttcagagaatcgttccattacaggaacactgtgactctggggaagataaacct 646
QY 481 gagggcagacgatacttgaaagtcattataaattggggaagaagaatgggtccatctt 540
Db 647 gaagccagacgatacttgaaagtcattataaattggggaagaagaatgggtccatctt 706
QY 541 cctgaacaagtacagaatgaaatcaaatcaatgaagaaagaatgagtgagctatgtatt 600
Db 707 ccnnagaagtacagaatganatcaaatcaatgaagaaagaatgagtgagctatgtatt 766

```

Qy	601	gattttaacaaaaaccctcaatgagatgataccttctctctgatttatttccaggctgaactt	650
Db	767	gactttaacaaaaaccctcaatnaggatgataccttccntgtgttttccaggctgagctt	826
Qy	661	ggtagctcttcctgagtgatttcattgacaggtttagaaaaagacagatgatgacaagtataa	720
Db	827	ggtagctcttcctgagtgatttcattgacaggtttagaaaaagtgaggatgacaaagtataa	886
Qy	721	attacctttaaaatataccacactatttccctgctcatgaagaataattgtatccctgaaac	780
Db	887	attacctttaaaatataccacactatttccctgctcatgaagaataattgtatttctctgaaac	946
Qy	781	agaagaagatggaattggcttttaatacaaggtcgaagaggaaaaacacataatttgg	840
Db	947	agaaggagaatggaattggcttttaatacaagatgcaaaaggnaaacccgcattctg	1006
Qy	841	caagcagctactcccactgcgaaccaaagggtggccaaactactcgttatagcacacatgct	900
Db	1007	cagcagctactcccactgcgggccaaagtggccaaactcctgggctatagcacacatgct	1066
Qy	901	gacttcgtcttgaattgaacactgcaagagcacaagccgcTaaacagcyttcttagat	960
Db	1067	gacttcgtcttcgagatgaacactgcaagagcacaagccgcgTaaacagcgttctttagac	1126
Qy	961	gatttaagccagaagttaaaacccctgggtgaagcagaacgagaggttatttgaatttg	1020
Db	1127	gatttaagccagaataataaaacccgtgggtgaggcagaacgaggttattcttgagtttg	1186
Qy	1021	agaaaaaaggaatgcaagacacaggggttttgaatatgatgggaaaaatcaatgctcgggt	1080
Db	1187	agaaaaaaggaatgcaagagagaaggtttgagtatgacgggaaaaatcaatgctcggac	1246
Qy	1081	ctatatctactcagctcagacagagaactcaagtattccatagaccagaaggttcctc	1140
Db	1247	ctcatctactcagctcagacagagaactcaagtattccatagaccaggaggttcctc	1306
Qy	1141	aaggaaatacttcccattgaggtggtcactgaa-----ggcttctggaacacctaccag	1194
Db	1307	aaggagtaacttcccccnntnagggtggctcancnaggggcctgctgaaactctaccag	1366
Qy	1195	gagttgtgggacttttcatttgaaacaaatgacagatgctcatgttttggaacaaagagtgtt	1254
Db	1367	gagttgtgggacttttattcgagcaagtggccagcgcatactgatttggaacccagaagtc	1426
Qy	1255	acactttactgtgaagataaaagctacagagagaagtattgggacaggttctctattggac	1314
Db	1427	acactttacactgtgaaggacaaggtctacagagaagtgcctggggcaggttctacctggac	1486
Qy	1315	ctctatccaaaggaaagaaaaatacaatcatcgccctgcttccgctccagcctggtctgc	1374
Db	1487	ctctatccngggaaagaaaaataaatcatgcagnctgcttggctccagcctggctgc	1546
Qy	1375	ctctctgcttgaaggacccgagatgagcagtggtgcctctgctgggtgaacttctcacag	1434
Db	1547	ctgctnccogatgggacggcagatgctgctgtggtgcctctggttgaactnnccag	1606
Qy	1435	ccagtgagcaggtcgctccctctctctctgagacacacagaggtgaggaccttacttccatg	1494
Db	1607	ccagtgagcggccgcnccctctctctgagggcgagatgaagtgmgncctacnccccatgan	1666
Qy	1495	tttggttcacgtgatgcatcagatttgcacagactgatttgcacgatttagcgaaaca	1554
Db	1667	tttggttcacgtgatgcatcagcttctgtgcncacagactgatttgcacgatttagtgaaca	1726
Qy	1555	aatgtggaactgaactttgtagagdgccatcgcaaatgcttgaataattgggtgtggggac	1614
Db	1727	aatgtggaactgaactttgtagagdgcttcacaaatgcttgagaactgggtgtggggac	1786
Qy	1615	gtcgattccctccgaagattgtcaaaacattataaagatgaagccctatttcagacagat	1674
Db	1787	atcgatcctctccgangatctnnaaaacattataaagatggnnaccctatgacagatgat	1846
Qy	1675	ctgcttgaaaaactgtgtcttcttagctgtgctcaacacaggtcttctgacccctgcgccag	1734

### RESULTS

РЕЗУЛТАТИ  
АДАН14379

AAH14373  
ID AAH14379 standard: cDNA: 1658 BP.

AAH14379:

26-JUN-2001 (first entry)

Human cDNA sequence SEQ ID NO:11796.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

28-JUL-2000: 2000EP-0116126.

PR 29-JUL-1999: 99JP-0248036.

PR 11-JAN-2000: 2000JP-0118776.

PR 02-MAY-2000; 2000F-0183787  
PR 09-JUN-2000; 2000JP-0241899

XX PA (HELJ-) HELIX RES INST.

XX  
PT  
Ota T  
Tsuqai T  
Nishikawa

PL 1511 S, Sugiyama T,  
vv

DR WPI; 2001-318/49/34.  
yy

Primer sets for synthesizing polynucleotides, particularly

PT and/or diagnosis of

[illegible][illegible]

CC full-length cDNAs defined in the specification. Where a primer set

CC to the complementary strand of a polynucleotide which comprises one

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combina-

complementary strand of a polynucleotide which comprises a 5'-end





DR P-PSDB; AAR24259.

XX Human amyloidin protease - used for cleaving Met-Asp bond in  
PT amyloid-like substrate for identifying protease inhibitors

XX PS Claim 37; Page 53-54; 62pp; English.

XX CC The complete sequence of the human amyloidin gene was obtained by  
CC sequencing two clones: clone CHL57 (codons 1-480) and clone CHL53  
CC (codon 56 to STOP codon). The clones were isolated after screening cDNA  
CC libraries prepared from the following sources: normal human brain,  
CC superior temporal gyrus and human embryonic kidney cell line 293.  
CC Examination of the nucleotide sequence reveals that there is no  
CC clear hydropobic leader sequence. The deduced amino acid sequence  
CC of the amyloidin protease contains the motif X-H-E-F-G-H-X at the  
CC active site typical of zinc-dependent metalloproteinases; the two  
CC histidine residues coordinate the Zn (2+) in the active centre and  
CC the glutamate is involved in bond-breaking.  
CC See also AAR24260-3, AAR24266-7 and AAQ24876-Q24887.

XX SQ Sequence 2070 BP; 434 A; 626 C; 578 G; 332 T; 0 other;

Query Match 37.4%; Score 790.8; DB 13; Length 2070;  
Best Local Similarity 62.4%; Pred. No. 3.7e-219;  
Matches 1236; Conservative 1; Mismatches 743; Indels 0; Gaps 0;

QY 128 ctggcagaaatgttttaagatgggattcttccacagagcaaatataaaacagaactgagg 187  
DB 56 ctgttgtaaacagactcgtgtggacctgagtcgccacagatagagagcgccacagg 115  
QY 188 agctcattgtcagacaaacaggtgtacgatgctgttggaatgcctgattgaggag 247  
DB 116 agctcatcgacagacaaacagcgctgtatgaccaggttggcaccagagattgaggag 175  
QY 248 taacttacgagaactgtctgcaggcaactggcagatgtagaagtaagtatatatggaaa 307  
DB 176 tgcctacagagacgcgtcgaagcgctggcagatgtgaggtcacctacacagttcaga 235  
QY 308 ggaccatctagactttcccagactgtatcctctgacaaagaagtagcagcaacagta 367  
DB 236 ggaatattcttgacttcccagactgttcccctccaaagacatccggacagcagca 295  
QY 368 cagaagcagacaaagactttctgtttgattgattgagatgagatgagagagatatat 427  
DB 296 cagagccgacagaagctctctgattcgcgtggagatgagatgagagagacgtgt 355  
QY 428 ttgagagaattgttcattacagaaaacctgtatctgtgggaagataaaacctgaggcca 487  
DB 356 accagagatcgtgtggtccaggagaaagtctcagaaggaactcactgaggccgaggtg 415  
QY 488 gacgatacttgaaaaagtcattataaaatgggaaaaagaataatgggtccacttcttgaa 547  
DB 416 cgcgttacttgagcggcgtatcaagctggcggcggagaaatgggttccactcccccag 475  
QY 548 aagtcagaaatgaatacaatcaatgaagaaaagaatgagtgcgtatgtattgatttta 607  
DB 476 agactcagaaaaacatacaacgcatacaagaagaagtgagccttctgtgcatgacttca 535  
QY 608 acaaaaacctcaatgagatgataaccttcttctgtattttccaaggtgaaacttgggtc 667  
DB 536 acaagaacctgaacgaggacacgaccttcttcccttcacgtccacgtcagagtaggggc 595  
QY 668 ttctgattgatttcatgacagtttagaaaagacagatgacaaagtataaaattact 727  
DB 596 tcccagagactttctgaactccctcctggagaagatggagacggaagttgaaggtcacccc 655  
QY 728 taaaatacccacactatttccctgtcatgaagaaaatgttgtatccctgaaaccagagaa 787  
DB 656 tcaagtacccccattacttccctcctctgaagaaatgccagctgacctgagaccagagga 715  
QY 788 ggaatggaaatggcttttaatacagaaggtgcaagagggaacacacataattttgcagcag 847  
DB 1796 gccatctggcagggtgggtacacgaccccgactactcgggtacactggtggagggcgtatt 1855

DB 716 aagtggaggaggccttcaactgcgggtgcaaggaggagaaactgcgctatcctcaaggagc 775  
QY 848 tactcccactgcgaacaaagtgccaaactactcgggttatagcacacatgctgacttcg 907  
DB 776 tggcgacgtcgcgggcgcagaaagtcgcgcctcgtcgtgggttccacacgcacgcgactatg 835  
QY 908 tccttgaatgaacactgcaaaagacacaaagccgctaacacagcytttctagatgatttaa 967  
DB 836 tcctggagatgaacatggcgaagaccagccagaccgtggccaccttctagatgagctgg 895  
QY 968 gccagaagttaaacaccttgggtggaacagaaacagagatttatattgaattgaagaaaa 1027  
DB 896 cgcagaagctgaagcccttggggagcgtgcggtgattcttggagctgaagcgtg 955  
QY 1028 aggaatcaaaagacaggggttttgaatatgatggaaaatcaatgcctgggatctatt 1087  
DB 956 cggagtcgagcggcggggctgccttcgacggcgcgcctcgtggacatgcgtgacctggacatgcgct 1015  
QY 1088 actacatgactcagacagagaaactcaagtattccatgacacaaagagttcctcaaggaa 1147  
DB 1016 actacatgaaccaggtgagagagacgcgctactcgtggaccgaacactgctcaaggagt 1075  
QY 1148 acttccaaattgaggtgctactgaagcttgcgtgaacacatccacagagattgttgggac 1207  
DB 1076 acttcccctgaggtggtcagcgcacggctgctgggcacatctacagagactcctggggc 1135  
QY 1208 ttctattgaacaaatgacagatgctcatgttttgaaacaaagagtttacctttatactg 1267  
DB 1136 tggccttcacacagagagggcgccagtcgctggcatgagagagtgcggtctacacgg 1195  
QY 1268 tgaaggataaaagctacagggagaagtatttggacagcttctatttggacctctatccaaagg 1327  
DB 1196 cgaggcagcggcctcggggagggtggtcggaagtcttacctggaacctgtaccgcggg 1255  
QY 1328 aagaaaaatacaatcatcagcggcctgttcgtccacagctgctccttctgctgctg 1387  
DB 1256 aagaaaaatgcggcagcggcctgttggcctgcagcccggtcgtcgtcggcaggtg 1315  
QY 1388 gaagccggatgatggcagtggtgcctcgtggtgaacttctcagaccagtggtggcaggtc 1447  
DB 1316 ggagccagatgcctatcgggccatcggtggcgaacttcaacaaagccacagccgagc 1375  
QY 1448 gtccctctctctgagacacagcaggtgaggaacttacttcatgagtttgggtcaacgtga 1507  
DB 1376 cgccctcgtcgtcagcatgacgaggtggagacctacttccatgagtttggcagctga 1435  
QY 1508 tgcacgattgtgcacagactaattttgcagatttagcggaacaaatgtgaaactg 1567  
DB 1436 tgcaccagctcgtcccgagggttcgccaatgttcagcgggacccacgtggagcggg 1495  
QY 1568 actttagaggtgccatcgcaaatgcttgaataatgggtgtgggacgtcgtatccctcc 1627  
DB 1496 actttagagggccgctcgagatgctggagaactgggtgtgggagcagagccgctgc 1555  
QY 1628 gaagattgtcaaaacattataaagatggaagccctatttgcagacgatctgttgaanaac 1687  
DB 1556 tgcggatgtcggcaactaccgcacagcagcgcgtgccccgggagctcctggagaagc 1615  
QY 1688 ttgtgtcttaggtgtcaacacaggtcttgcacctgctgcagagattgttttgagca 1747  
DB 1616 tcatgtgctccggcagggccacacagcctcttcagcctgcgcagatcgtcctcgcca 1675  
QY 1748 aagtgtgactctcttcatcacaacacatcgctggatgctgcaagtgaatgccaat 1807  
DB 1676 aggtggaccagccctgcacgcagacgacgacagcccgccgaggtatgctgcggcg 1735  
QY 1808 actgctcagaaaattattaggagtttgcagctactccagggcacaataatgacagactttg 1867  
DB 1736 tctgcagagagatcctcgtgggtcccgccagcgcaggaacacacatgctgcaaccttcg 1795  
QY 1868 gacatttggcaggggatacagatggcaaatatttgatatcttttggagtgaaatattt 1927  
DB 1796 gccatctggcagggtgggtacacgaccccgactactcgggtacactggtggagggcgtatt 1855



```
PD XX 06-APR-2000.
PF XX 23-SEP-1999; 99WO-US22226.
PR XX 28-SEP-1998; 98US-0102161.
PR XX 28-SEP-1998; 98US-0102180.
PR XX 29-SEP-1998; 98US-0102380.
PR XX 08-OCT-1998; 98US-0103815.
PR XX 27-OCT-1998; 98US-0105877.
PA XX (CHIR ) CHIRON CORP.
PA XX (HYSE-) HYSEQ INC.
XX XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
XX XX Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
XX XX Lamson G, Drmanac K, Drmanac R, Dickson M, Drmanac S, Labat I;
XX XX Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX XX WPI; 2000-293155/25.
XX XX Polynucleotide library comprising 1079 defined sequences, useful in
XX XX the form of an array to detect cancer or susceptibility to cancer -
XX XX
XX XX Claim 1; Page 464; 502pp; English.
XX XX
XX XX The present invention describes a library of polynucleotides comprising
XX XX 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described
XX XX are: (1) an isolated polynucleotide (I) having at least 90% identity to
XX XX one of the 1079 sequences; (2) a recombinant host cell containing (1);
XX XX (3) an isolated polypeptide (II) encoded by (1); (4) an antibody that
XX XX specifically binds to (II); (5) a vector comprising (1); and (6) a method
XX XX of detecting differentially expressed genes correlated with a cancerous
XX XX state of a mammalian cell comprising detecting a gene product encoded by
XX XX one of the 1079 sequences given in the specification. The polynucleotides
XX XX are used to monitor patients having (or susceptible) to cancer to detect
XX XX potentially malignant events at a molecular level before they are
XX XX detectable at a gross morphological level. The polynucleotides are also
XX XX useful for monitoring the efficacy of various therapies and preventive
XX XX interventions. Polynucleotide probes based on the disclosed sequences
XX XX are useful for chromosome mapping and detection of transcription levels.
XX XX The 1079 polynucleotide sequences were derived from a human colon cancer
XX XX cell line Kml2L4-A cDNA library.
XX XX
XX XX Sequence 339 BP; 97 A; 72 C; 81 G; 89 T; 0 other;

Query Match 16.0%; Score 339; DB 21; Length 339;
Best Local Similarity 100.0%; Pred. No. 2.2e-88;
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1548 cggacaataatggaactgactttgttagagtgccatcgcaaatgctgaaattgggt 1607
Db 1 cggacaataatggaactgactttgttagagtgccatcgcaaatgctgaaattgggt 60
Qy 1608 gtggagcgtgattccctccgaagattgtcaaaacattataaagatggaagccctattgc 1667
Db 61 gtggagcgtgattccctccgaagattgtcaaaacattataaagatggaagccctattgc 120
Qy 1668 agacgactgtgtgaaacattgttcttcttaggtgtgcaacacaggtctcttgaccct 1727
Db 121 agacgactgtgtgaaacattgttcttcttaggtgtgcaacacaggtctcttgaccct 180
Qy 1728 gcgccagattttttgacaaagtgtgatcagctctcttccatcccaacacatcgctggatgc 1787
Db 181 gcgccagattttttgacaaagtgtgatcagctctcttccatcccaacacatcgctggatgc 240
Qy 1788 tgcagtgaaatgcaaaatactgctcagaataattagagttgagctactccaggcac 1847
Db 241 tgcagtgaaatgcaaaatactgctcagaataattagagttgagctactccaggcac 300
Qy 1848 aaatgcccagctacaccttttgacatttggcagggggata 1886
Db 301 aaatgcccagctacaccttttgacatttggcagggggata 339
```

```
RESULT 7
AAC98797
ID AAC98797 standard; cDNA; 900 BP.
XX
AC AAC98797;
XX
XX 09-MAR-2001 (first entry)
XX
XX Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:25.
XX
XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
XX detection; diagnosis; identification; cytostatic; neuroprotective;
XX nontropic; immunomodulatory; relaxant; contraceptive; gynaecological;
XX antiinflammatory; cardiant; gene therapy; chromosome mapping;
XX linkage analysis; tissue identification; tissue typing; forensic;
XX neural; immune system; muscular; reproductive; gastrointestinal;
XX pulmonary; cardiovascular; renal; proliferative; ss.
XX
XX Homo sapiens.
XX
XX WO200055320-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05989.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI; 2000-579444/54.
XX
XX P-PSDB; AAB54032.
XX
XX New nucleic acid that is a pancreatic cancer antigen for preventing,
XX treating, or ameliorating a medical condition, particular pancreatic
XX cancer, or for use in assays for diagnosing a pathological condition -
XX
XX Claim 1; Page 512-513; 1379pp; English.
XX
XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
XX proteins, called pancreatic cancer antigens, given in AAB54008 to
XX AAB54466. The human pancreatic cancer antigens have cytostatic,
XX neuroprotective, nontropic, immunomodulatory, relaxant, contraceptive,
XX gynaecological, cardiant and antiinflammatory activities, and can be used
XX in gene therapy. The polynucleotide and proteins can be used for
XX preventing, treating, or ameliorating a medical condition or in assays
XX for diagnosing a pathological condition or a susceptibility to one in a
XX subject. Binding partners to the proteins and the activity of the
XX proteins can be identified. The pancreatic cancer antigens can be used to
XX detect, treat or prevent pancreatic disorders, especially cancer.
XX Agonists and antagonists to the antigens can be screened for. The
XX pancreatic cancer antigen polynucleotides can be used to design nucleic
XX acid hybridisation probes that can be used in chromosome mapping, linkage
XX analysis, tissue identification and/or typing and a variety of forensic
XX analysis, tissue identification and/or typing and a variety of antibodies
XX which are used to purify, detect and target the polypeptides, including
XX both in vivo and in vitro diagnostic and therapeutic methods. The
XX proteins can be used to treat or prevent neural, immune system, muscular,
XX reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
XX proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
XX sequences used in the exemplification of the present invention.
XX
XX Sequence 900 BP; 160 A; 284 C; 299 G; 147 T; 10 other;

Query Match 14.6%; Score 308.8; DB 21; Length 900;
Best Local Similarity 64.1%; Pred. No. 2.3e-79;
Matches 480; Conservative 0; Mismatches 268; Indels 1; Gaps 1;
```



















GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: January 14, 2002, 18:38:36 ; Search time 72.17 Seconds

(without alignments)  
722.566 Million cell updates/sec

Title: US-09-833-782-2

Perfect score: 3668

Sequence: 1 MIARCLLAVLSRLRVGSGRI.....FLKREPQKAFMLMSRGLHAP 704

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

A.Geneseq\_1101.\*  
1: /SID52/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3668	100.0	704	22	Human aminopeptidase
2	2979	81.2	709	13	Rabbit endopeptidase
3	2331	63.5	689	13	Human amyloidin pr
4	1981	54.0	381	22	Human protein sequ
5	702	19.1	678	21	N. gonorrhoeae amin
6	702	19.1	678	21	N. meningitidis am
7	702	19.1	678	21	N. gonorrhoeae amin
8	702	19.1	678	21	N. meningitidis am
9	702	19.1	678	21	Neisseria gonorrhoe
10	702	19.1	678	21	Neisseria meningit
11	702	19.1	678	21	Neisseria gonorrhoe

12	702	19.1	678	21	AA1975802	Neisseria meningit
13	694	18.9	678	21	AA1958576	N. meningitidis am
14	694	18.9	678	21	AA1958579	N. meningitidis am
15	694	18.9	678	21	AA1925646	N. meningitidis am
16	694	18.9	678	21	AA19474506	Neisseria meningit
17	694	18.9	678	21	AA19474509	Neisseria meningit
18	694	18.9	678	21	AA1975801	Neisseria meningit
19	694	18.9	678	21	AA1975804	Neisseria meningit
20	691	18.8	678	21	AA1925649	N. meningitidis am
21	659.5	18.0	783	21	AA1951467	Arabidopsis thalia
22	659.5	18.0	785	21	AA1951466	Arabidopsis thalia
23	659.5	18.0	793	21	AA1951465	Arabidopsis thalia
24	583	15.9	678	22	AA192271	C glutamic prote
25	567	15.5	211	21	AA194032	Human pancreatic c
26	480.5	13.1	476	21	AA1958574	N. meningitidis am
27	480.5	13.1	476	21	AA1925644	N. meningitidis am
28	475	12.9	491	21	AA1958578	N. gonorrhoeae amin
29	475	12.9	491	21	AA1925648	N. gonorrhoeae amin
30	475	12.9	491	21	AA194507	Neisseria gonorrhoe
31	475	12.9	491	21	AA1975803	Neisseria gonorrhoe
32	461.5	12.6	344	21	AA194505	Neisseria meningit
33	461.5	12.6	344	21	AA1975799	Neisseria meningit
34	446.5	12.2	555	22	AA1979072	Corynebacterium gl
35	381	10.4	355	22	AA1979073	Corynebacterium gl
36	270	7.4	234	11	AA1940411	Protein product of
37	169	4.6	41	22	AA1918211	Peptide #4645 enco
38	169	4.6	41	22	AA1930705	Peptide #4742 enco
39	169	4.6	41	22	AA1905830	Peptide #4512 enco
40	146	4.0	29	21	AA1953253	Human type enzyme
41	133	3.6	1392	20	AA1906999	Restin protein seq
42	132.5	3.6	1294	22	AA1941110	Human polypeptide
43	131	3.6	1427	12	AA190534	Human 160kd mediat
44	128.5	3.5	2048	22	AA1940027	Human polypeptide
45	127.5	3.5	1700	22	AA1939324	Human polypeptide

## ALIGNMENTS

RESULT 1

AA1973485	AA1973485 standard; Protein; 704 AA.
ID	AA1973485 standard; Protein; 704 AA.
XX	
AC	AA1973485;
XX	
DT	25-JUL-2001 (first entry)
XX	
DE	Human aminopeptidase 22196.
XX	
KW	Human; aminopeptidase 22196; neutral zinc metalloproteinase;
KW	metalloenzyme; analgesic; cytostatic; lung cancer; colon cancer; tumour;
KW	pain; drug screening; spleen disorder; infectious disease;
KW	immune disorder; lung disorder; colon disorder; liver disorder;
KW	uterine disorder; brain disorder; skin disorder; cardiac disorder;
KW	vascular disorder; bone disorder; gene therapy.
XX	
OS	Homo sapiens.
XX	
Key	Location/Qualifiers
FT	Modified-site 11..13
FT	/note= "Ser is O-phosphorylated by protein kinase C"
FT	Modified-site 59..62
FT	/note= "Thr is O-phosphorylated by casein kinase II"
FT	Modified-site 104..107
FT	/note= "Thr is O-phosphorylated by casein kinase II"
FT	Modified-site 114..116
FT	/note= "Ser can be O-phosphorylated by protein kinase C or by casein kinase II"
FT	Modified-site 123..126
FT	/note= "Thr is O-phosphorylated by casein kinase II"
FT	Modified-site 127..130
FT	/note= "Ser is O-phosphorylated by cAMP/cGMP-dependent protein kinase"
FT	

FT Modified-site 130..133 /note= "Ser is O-phosphorylated by casein kinase II"  
FT Modified-site 137..139 /note= "Ser is O-phosphorylated by protein kinase C"  
FT Modified-site 157..165 /note= "Tyr is O-phosphorylated by a tyrosine kinase"  
FT Modified-site 169..171 /note= "Ser is O-phosphorylated by protein kinase C"  
FT Modified-site 172..175 /note= "Amidation site"  
FT Modified-site 190..192 /note= "Ser is O-phosphorylated by protein kinase C"  
FT Modified-site 193..196 /note= "Ser is O-phosphorylated by cAMP/cGMP-dependent protein kinase"  
FT Modified-site 216..219 /note= "Ser is O-phosphorylated by casein kinase II"  
FT Modified-site 233..239 /note= "Tyr is O-phosphorylated by a tyrosine kinase"  
FT Modified-site 234..237 /note= "Thr is O-phosphorylated by casein kinase II"  
FT Modified-site 242..244 /note= "Thr is O-phosphorylated by protein kinase C"  
FT Modified-site 260..262 /note= "Thr is O-phosphorylated by protein kinase C"  
FT Modified-site 298..301 /note= "Thr is O-phosphorylated by casein kinase II"  
FT Modified-site 308..310 /note= "Thr is O-phosphorylated by protein kinase C"  
FT Modified-site 312..314 /note= "Thr is O-phosphorylated by protein kinase C"  
FT Modified-site 323..325 /note= "Ser is O-phosphorylated by protein kinase C"  
FT Modified-site 366..369 /note= "Thr is O-phosphorylated by casein kinase II"  
FT Modified-site 392..397 /note= "Gly is N-myristoylated"  
FT Modified-site 396..399 /note= "Thr is O-phosphorylated by casein kinase II"  
FT Modified-site 415..418 /note= "Asn is N-glycosylated"  
FT Modified-site 422..425 /note= "Thr can be O-phosphorylated by protein kinase C or by casein kinase II"  
FT Modified-site 453..458 /note= "Gly is N-myristoylated"  
FT Modified-site 475..478 /note= "Asn is N-glycosylated"  
FT Modified-site 488..495 /note= "Tyr is O-phosphorylated by a tyrosine kinase"  
FT Modified-site 518..521 /note= "Thr is O-phosphorylated by casein kinase II"  
FT Modified-site 541..543 /note= "Ser is O-phosphorylated by protein kinase C"  
FT Modified-site 543..546 /note= "Ser is O-phosphorylated by cAMP/cGMP-dependent protein kinase"  
FT Modified-site 552..557 /note= "Gly is N-myristoylated"  
FT Modified-site 575..577 /note= "Thr is O-phosphorylated by protein kinase C"  
FT Modified-site 582..585 /note= "Ser is O-phosphorylated by casein kinase II"  
FT Modified-site 591..594 /note= "Asn is N-glycosylated"  
FT Modified-site 592..595 /note= "Thr is O-phosphorylated by casein kinase II"  
FT Modified-site 627..632 /note= "Gly is N-myristoylated"  
FT Modified-site 674..679 /note= "Gly is N-myristoylated"  
XX PN WO200123590-A2.

XX 05-APR-2001.  
XX PD 02-OCT-2000; 2000WO-US27214.  
XX PF 30-SEP-1999; 99US-0409180.  
XX PR (MILL)-MILLENNIUM PHARM INC.  
XX PA Kapeller-Libermann R, White D, Silos-Santiago I;  
XX PI WPI; 2001-300101/31.  
XX DR N-PSDB; AAF84499.  
XX PT Novel human aminopeptidase, used to treat colon and lung cancer, and to  
XX identify modulators used to treat e.g. cirrhosis  
XX PS Claim 1; Fig 1A-B; 109pp; English.  
XX CC This sequence represents a novel human aminopeptidase, designated  
CC 22196 cDNA clones encoding human aminopeptidase 22196  
CC were identified in a bone marrow cDNA library using an EST (expressed  
CC sequence tag) selected on the basis of homology to aminopeptidase  
CC sequences, and were assembled to give the full-length cDNA.  
CC Aminopeptidase 22196 is thought to be a metalloenzyme, as it contains  
CC a zinc-binding region signature found in neutral zinc metalloproteinases.  
CC Expression analysis indicated that the aminopeptidase is highly  
CC expressed in normal osteoblasts, testis, skeletal muscle, foetal kidney  
CC and foetal liver. It is also expressed in normal breast, lung, and colon  
CC tissue, and is overexpressed in lung and colon cancer. Expression was  
CC found to be downregulated during stromal cell osteoblast lineage  
CC maturation. Aminopeptidase 22196 may be used to treat pain, and cancers  
CC of the lung and colon. It can also be used in drug screening to identify  
CC compounds which bind to it and modulate its activity. Such compounds may  
CC be used in the treatment of a wide variety of conditions, including  
CC spleen disorders (e.g., splenomegaly); infectious diseases (e.g.,  
CC malaria, tuberculosis); immune disorders (e.g., rheumatoid arthritis,  
CC systemic lupus erythematosus, acquired immunodeficiency syndrome (AIDS));  
CC lung disorders (e.g., pulmonary embolism, emphysema, bronchial asthma);  
CC colon disorders (e.g., diarrhoea, Crohn's disease, ulcerative colitis);  
CC liver disorders (e.g., hepatic cirrhosis, hepatitis); uterine disorders  
CC (e.g., endometriosis); brain disorders (e.g., meningitis, Alzheimer's  
CC disease, Parkinson's disease, Huntington's disease); skin disorders  
CC (e.g., vitiligo, urticaria, skin tumours, psoriasis); cardiac and  
CC vascular disorders (e.g., myocardial infarction, chronic ischaemic heart  
CC disease, hypertension, atherosclerosis); and bone disorders (e.g.,  
CC osteoporosis, Paget's disease). Nucleotides encoding aminopeptidase 22196  
CC may be used in gene therapy.  
XX SQ Sequence 704 AA;  
Query Match 100.0%; Score 3668; DB 22; Length 704;  
Best Local Similarity 100.0%; Pred. No. 9.5e-308;  
Matches 704; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 M I A R C L L A V R S L R V G G S R I L L R M T L G R V M S P L Q A M S S Y T V A G R N V L R W D L S P E Q I K T R 60  
Db 1 m i a r c l l a v r s l r v g g s r i l l r m t l g r v m s p l q a m s s y t v a g r n v l r w d l s p e q i k t r 60  
QY 61 T E E L I V Q T K Q V D A V G M L G I E E V T Y E N C L Q A L A D V E K Y I V E R T M L D F P Q H V S S D K E V R A 120  
Db 61 t e e l i v q t k v d a v g m l g i e e v t y e n c l q a l a d v e k y i v e r t m l d f p q h v s s d k e v r a 120  
QY 121 A S T E A D K R L S R F D E M S M R G D I F E R I V H L Q E T C D L G K I K P E A R Y L E K S I K W G R N G L H L 180  
Db 121 a s t e a d k r l s r f d e m s m r g d i f e r i v h l q e t c d l g k i k p e a r y l e k s i k m g r n g l h l 180  
QY 181 P E Q V Q N E T K S M K K R M S E L C I D F N K N L N E D D F L V F S K A E L G A L P D D F T D S L E K T D D D K Y K 240  
Db 181 p e q v q n e l k s m k k r m s e l c i d f n k n l n e d d t f l v f s k a e l g a l p d d f l d s l e k t d d d k y k 240  
QY 241 I T L K Y P H Y F P V M K K C C I P E T R R R M E M A F N T R C K B E N T I I L Q O L L P L R T K V A K L I G Y S T H A 300

|||||  
Db 241 itlkyphfvmkkccipetrrrmemafrnckeeentilqlplrtkvakllystha 300  
QY 301 DFVLEMTAKSTSRVTAFLDDLSOKLPLGEAREFELNKKKECKDRGFEDGKINAWD 360  
Db 301 dfvlemtakstsrvtaflddlsqklplgeaerefilnkkeckdrgfeydggkinawd 360  
QY 361 LYYMTQTTEELKYSIDQEFLEKVFPEVTEGLLNTYQELLGLSPQMTDAHWNKSVTL 420  
Db 361 lyymtqtteelkysidqeflekyfpievteglntyqellglsefemtahwnksvtl 420  
QY 421 YTVKDKATGEVLGGFYLDLYPREGKYNHAAACFGLOPGCLLPDGSRRMAVAALVNFSPQV 480  
Db 421 ytvkdkatgevlgqfyldlypregkynhaacfglpgcllpdgsrrmavaalvvnfsqpv 480  
QY 481 AGRPSLLRHDEVRTYFHFHGHVHMHOICAOQDFARFSGTNVETDFVEVPSQMLENWWVDY 540  
Db 481 agrpsllrhdevrtyfhefghvhmhqcactdfarfsgtnvetdfvevpsqmlenwwvdv 540  
QY 541 SLRRLSKHYKDGSPADDDLEKLVASRLVNTGLTLRQIVLSKVQDSLHTNTSLDAASEY 600  
Db 541 slrrlskhykdgspiadddlleklvasrlvntgltlrqivlskvdqslhtntsldaasey 600  
QY 601 AKYCSILGVAATPGTNMPATFGHLAGGYDGYGYLWSEVFSMDMFYSCFKKEGIMNPE 660  
Db 601 akycseilgvaatpgtnmpatfghlaggydgygyylwsevfsmmfycfkkegimnpe 660  
QY 661 VGMKYRNLLKGGSLDGMDLHNFILKREPNOKAFILMSRGLHAP 704  
Db 661 vgmkyrnllkpggsldgmdlmhnlkrepnqkfllmsrglhap 704

RESULT 2  
AAR26114  
ID AAR26114 standard; Protein; 709 AA.  
XX AC AAR26114;  
XX DT 03-FEB-1993 (first entry)  
XX DE Rabbit endopeptidase (preliminary sequence).  
XX KW microsome; metalloproteinase; polymerase chain reaction.  
XX OS Oryctolagus cuniculus.  
XX PN WO9213080-A.  
XX PD 06-AUG-1992.  
XX PF 28-JAN-1992; 92WO-US00731.  
XX PR 28-JAN-1991; 91US-0646997.  
XX (UNIW ) UNIV WASHINGTON.  
XX PA Davie EW, Kawabata S;  
XX PI WPI; 1992-284665/34.  
XX DR N-PSDB; AAQ27307.  
XX Microsomal endo-peptidase isolated from rabbit liver - used to  
XX cleave protein substrates in-vivo and in-vitro  
XX Example 3; Fig 5; 48pp; English.  
XX The rabbit endopeptidase was purified from rabbit liver and  
XX digested with CNBr. Families of degenerate primers were designed  
XX based on the amino acid sequences of the CNBr fragments. The  
XX primers were used to amplify sequences from a rabbit cDNA library.  
XX An amplified fragment was isolated and used to probe the same  
XX library. Two overlapping clones produced a 2148bp sequence. A  
XX preliminary sequence analysis was performed in which much of the

CC 5' and 3'-most sequences are unassigned; the actual amino and  
CC carboxyl terminals of the protein have not yet been determined.  
CC The "Others" in the sequence correspond to codons in which 1 or more  
CC of the bases has not yet been identified.  
CC See AAR26107-R26113 and AAQ27305-Q27306.  
XX SQ Sequence 709 AA;  
Query Match 81.2%; Score 2979; DB 13; Length 709;  
Best Local Similarity 86.7%; Pred. No. 2.8e-248;  
Matches 568; Conservative 29; Mismatches 56; Indels 2; Gaps 1;  
QY 2 IARCLLAVSLRRVGRSRIILLRMTLGRVMSPLQAMSSYTVAGRNVLRLWDLSPEQIKTRT 61  
Db 54 iarcfsaxrghrvggrrllfkmrlgrvmsplqavssytaagrnlrvwldlseqiktrt 113  
QY 62 EELIVOTKQYDAVGMGLGIEEVTYENCALQADVEYKIVERTMLDFPQHVSDDKEVRAA 121  
Db 114 eeliaqtqkydsvgmldikdvtynclqaladvkyivertmldfpqhvstdevraa 173  
QY 122 STEADKRLSRFDIEMSRGDIPIFIVHLOETCDLGIKIPKPEARRYLEKSIKMGKRNGLHLP 181  
Db 174 steadkrlsrfdiemsmredifqirivhlqetcdlekikpearryleksvkmgrxglhlx 233  
QY 182 EOVSNEIKSMKMRMSLCLIDFNKNLNEDETFLVFSKAEALGALPDDFIDSLKTKDDDKYKI 241  
Db 234 xevqnxiksmkrmseicldfnknlnxddtxxvfskaelgaipddfidslkmdddkyki 293  
QY 242 TLKYPHYFPVMKKCCIPETRRMEMAFNTRCKEENTIILOQLLPLRTRTKVAKLLGYSTHAD 301  
Db 294 tlkyphyfpyvmkkccipetrrrmemafntrckextntvilqllplraqvakllgysthad 353  
QY 302 FVLEMTAKSTSRVTAFLDDLSOKLPLGEAREFELNKKKECKDRGFEDGKINAWDL 361  
Db 354 fvlemtakstsrvtaflddlsqklplgeaerefilnkkeckdrgfeydggkinawdl 413  
QY 362 VYVMTQTEELKYSIDQEFLEKVFPEV--VTEGLLNTYQELLGLSPQMTDAHWNKSVT- 419  
Db 414 hyymtqteelkysidqeflekyapxxxgxxrgllniyqellglxfeqvadanwnpxvt 473  
QY 420 LYTVDKATGEVLGGFYLDLYPREGKYNHAAACFGLOPGCLLPDGSRRMAVAALVNFSPQ 479  
Db 474 lytvdkatgevlgqfyldlypregkynhaxcfcglpgcclxpdgstrmlsaaalvvnxxqp 533  
QY 480 VAGRPSLLRHDEVRTYFHFHGHVHMHOICAOQDFARFSGTNVETDFVEVPSQMLENWWVDV 539  
Db 534 vagrpsllrhdevrtyfhefghvhmhqlcxqtdfarfsgtnvetdfvevpsqmlenwwvdi 593  
QY 540 DSLRRLSKHYKDGSPADDDLEKLVASRLVNTGLTLRQIVLSKVQDSLHTNTSLDAASE 599  
Db 594 dsrlrxxxkhykdxpxpiadddlleklvasrlvntgltlrqivlskvdqslhtntslldaase 653  
QY 600 YAKYCSILGVAATPGTNMPATFGHLAGGYDGYGYLWSEVFSMDMFYSCFKKE 654  
Db 654 yaryctdilgvaaxpgtnmpatfghlaggydgygyylwsexxsmdmxyscfkke 708  
RESULT 3  
AAR24259  
ID AAR24259 standard; Protein; 689 AA.  
XX AC AAR24259;  
XX DT 09-NOV-1992 (first entry)  
XX DE Human amyloidin protease.  
XX KW Alzheimer's disease; beta amyloid precursor protein; APP; zinc;  
XX metalloprotease; hAP; protease inhibitor; ss.  
XX OS Homo sapiens.

FH Key Location/Qualifiers  
 FT Active-site 472..478  
 FT /note= "His residues coordinate Zn2+ ion  
 FT and Glu is involved in bond-breaking"  
 FT Peptide 47..66  
 FT /note= "immunogenic"  
 XX W09207068-A.  
 XX 30-APR-1992.  
 XX 04-OCT-1991; 91WO-US07290.  
 XX 05-OCT-1990; 90US-0594122.  
 PR 30-SEP-1991; 91US-0766351.  
 XX (ATHE-) ATHENA NEUROSCIENCES INC.  
 PA (ELIL ) LILLY & CO ELI.  
 XX Dovey HF, Johnstone EM, Little SP, McConlogue L, Seubert PA;  
 PI Sinha S;  
 XX WPI; 1992-167148/20.  
 DR N-PSDB; AAO24875.  
 XX Human amyloidin protease - used for cleaving Met-Asp bond in  
 PT amyloid-like substrate for identifying protease inhibitors  
 XX Claim 3; Page 47-49; 62pp; English.  
 XX The deduced amino acid sequence of the amyloidin protease contains  
 CC the motif X-H-E-F-G-H-X at the active site typical of zinc-dependent  
 CC metalloproteinases. The enzyme hydrolyses the Met-Asp bond in an  
 CC amyloid-like substrate. e.g. the bond between Met 596 and Asp 597  
 CC of the beta-amyloid core peptide. Inhibitors of amyloidin activity  
 CC will be potentially useful in therapeutic intervention in Alzheimer's  
 CC Disease. See also AAR24260-3, AAR24266-7 and AAO24876-Q24887.  
 XX Sequence 689 AA;

Query Match 63.5%; Score 2331; DB 13; Length 689;  
 Best Local Similarity 64.9%; Pred. No. 2.3e-192;  
 Matches 426; Conservative 98; Mismatches 132; Indels 0; Gaps 0;

QY 46 NVLRWDLSPQIKFTTELLIVQTKQVDAVGMGLTEETVYENCLOALADVEVKYIVERTM 105  
 DB 22 nldrldlsaqleertlellqtkrvyqdvqtdqefedvsyestikaladvvtyvqrni 81  
 QY 106 LDFPOHVSDEVRRAASTADKRLSRFDIEMSRGDIPIRIVHLOETCDLGIKPEARRY 165  
 DB 82 ldfpohvpskdirtasteadkklsefovmsredyqrvlwlqekvqkdsilrpeaary 141  
 QY 166 LEKSTKMGKRNGLHLPEQVQNEIKSMKRMSELCDIFNKNLNEDDTFLVFSKAEALGALPD 225  
 DB 142 lerlklgrnglhlpretqenlkrkkllsldcfknknlnedttflptqlglgipe 201  
 QY 226 DFIDSLETKDDKYLTKYHPYPMKKCCIPETRRRMEAFNTRCKEENTIILOQLLP 285  
 DB 202 dflnskmedgklkvtklyphylpkkchvpetrrkveafncrckeaalkelvt 261  
 QY 286 LRTKVAKLGYSTHADFLVLENTAKSRVTAFLDLSQKLPLGEAREFRLNLKKEC 345  
 DB 262 lraqksrlglghthadyvlemmaktsgvatfidelakklpgeqeravilelakraec 321  
 QY 346 KDRGFYDGKTNANDLYYWTQTTELKYSIDQELKEYFPEVTVTEGLNTYQELLGLSF 405  
 DB 322 egrglpfgdgrawdmrymqveetyrcvqdqnlkkeyfpvqvvtghlglqellglaf 381  
 QY 406 EQMTDAHVNKSVTLTYVKDKATCEVLGQFVLDLYPREGKYNHAACFGLPGCLLPQGSR 465  
 DB 382 neegasawhedvrltyardasgevvqkfyldlypregkyghaacfglqpgclrqdgsr 441

QY 466 MMVAALVVFNSQPVAGRPSLLRHDEVTYFHERGHVHMQICAQDFARFSGTNVETDFV 525  
 DB 442 qiaiaamvanftkptadapsllqhddevetyfheghvnhqicsgaeafmgfsgthverdfv 501  
 QY 526 EVPSQMLENWWVDVSLRRLSKHYKDGSPDIADDLLEKLVASRLVNTGLTLRQIVLSKVD 585  
 DB 502 eapsgmlenwyweqepllrmarhyrtgsavprellekleliesrqantgifsrlqvlakvd 561  
 QY 586 QSLHTNTSLDAASEYAKYCEITGLVAAATPGTNMPATFGHLAGGYDGYGYVLMSEVFSMD 645  
 DB 562 qalhtqtdadpaeyarlcqeilgvpatpgtntpatfghlaggydaqyyglwsevyismd 621  
 QY 646 MFYSCFKKEGIMNPEVGMKYNRLILKPGSLDGMDFLHFLKREPNAQFLMSRGL 701  
 DB 622 mfhtrfkqevlnskvgmdyrscilrpggsedasamlrrfrlgrdpkqdafillskgl 677

## RESULT 4

AAB93028  
 ID AAB93028 standard; Protein; 381 AA.  
 XX AAB93028;  
 XX 26-JUN-2001 (first entry)  
 DT Human protein sequence SEQ ID NO:11797.  
 DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 KW Homo sapiens.  
 OS EPI074617-A2.  
 PN 07-FEB-2001.  
 PD 28-JUL-2000; 2000EP-0116126.  
 XX 29-AUG-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX (HELI-) HELIX RES INST.  
 PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX WPI; 2001-318749/34.  
 DR Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX Claim 8; SEQ ID 11797; 2537pp + CD ROM; English.  
 XX The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by







OS Neisseria gonorrhoeae.  
 PN WO200022430-A2.  
 PD 20-APR-2000.  
 XX 08-OCT-1999; 99WO-US23573.  
 XX 09-OCT-1998; 98US-0103794.  
 PR 30-APR-1999; 99US-0132068.  
 XX (CHIR ) CHIRON CORP.  
 PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
 PI Rappuoli R, Pizza M;  
 XX WPI; 2000-318079/27.  
 DR N-PSDB; AAA81285.  
 CC Isolated nucleotide sequences of Neisseria meningitidis which can be  
 CC used in the diagnosis and treatment of N. meningitidis infection and  
 CC other Neisserial infections, for example, N.gonorrhoea -  
 CC Claim 14; Page 93; 1760pp; English.  
 CC The present invention describes methods of obtaining immunogenic  
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414  
 CC represent specifically claimed Neisseria meningitidis genomic DNA  
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent  
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to  
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to  
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them, can be used in the manufacture of a composition. The  
 CC composition can be used as a medicament (or in the manufacture of a  
 CC medicament) for treating, preventing or diagnosing infection due to  
 CC Neisserial bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against Meningococcus B; against all serotypes;  
 CC and/or against all pathogenic Neisseriae. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.  
 XX SQ Sequence 678 AA;

Query Match 19.1%; Score 702; DB 21; Length 678;  
 Best Local Similarity 28.7%; Pred. No. 8.9e-52;  
 Matches 194; Conservative 127; Mismatches 305; Indels 50; Gaps 18;

QY 56 QIKTRTEELIVOTKQYDVGMLGTEVT----YENCLOALADVEKVIYIVERT----ML 106  
 DB 17 qiqtedikpavqt-aiaeargiaavkaqthtgwntverltg-----ltervgrigv 70  
 QY 107 DFPQHVSSDKVRAASTEADKRLSRFDIEMSRGDIIFERIVHLQETCDLGKIKPEARLYL 166  
 DB 71 shlnsvdtpelravynelpelvtftteigqdielynrftktiknspefatlsapaktkl 130  
 QY 167 EKSIRKNGKNGHLPEQVQNEIKSKMRMSELCIDFNKN-LNEDDTFLVF--SKAELGAL 223  
 DB 131 dhldrdvlsgealpperqaeiakiqtegaqisakfsqnvldatdafgiyfdadaaplagi 190  
 QY 224 PDD----FIDSLEKTDDDKYKITLXYPHYFPVWKCCIPETRRMEMAENTRCKE----- 274

DB 191 pedalamfaaaqsegkgykigqlqiphylavignrelreqiyrayvtraselndg 250  
 QY 275 --ENTIILOQLPLRTKVAKLIGYSTHADFLVEMNTAKSTSRVTAFLDLSOKRLPLGEA 332  
 DB 251 kfdntanidrtlenalktakllgfknyaelslatkmdtpeqvlnflhlarrakpyaek 310  
 QY 333 EREFILNLKKCKDGRGFEYDGKIN--AWDLVYYMTQTEELKYSIDOEFLKVFPTPEVT 390  
 DB 311 d-----laevkafarehlgldpqpwdlsyagcklreakyafsetevkkyfpgkvl 362  
 QY 391 EGLLNTYOELLGLSFQMDTAHWNKSVTLTYTKDKATGEVLGQFVLDLYPREGKYNHAA 450  
 DB 363 aglfaikklygigfaekt-vpwnhkdvryfelqq-ngktggvymdlyaregkrggaw 419  
 QY 451 CGLGQPCCLLPGRSMMVAALVWVNSQVAGRPSSLRHRDEVRTYFHERGHVHQICAQT 510  
 DB 420 mndykgrrrfadgtlqlptayivcnfappvggkearlshdeltilfhctghghlhltdv 479  
 QY 511 DFAPTSQTN-VETDFEVPQMLENWNWVDVSLRRLSKYKDGSPITADDLLEKLVASRLV 569  
 DB 480 delgvsngingvewdaveipsgfmenfvweynvlagmsaheetgeplpkelfdkmlaaknf 539  
 QY 570 NPLTLTLRLQIVLSKVDQSLHTNTSLDAASEYAKYCYSEI---LGVAAATPGTNMPA-TFGHL 625  
 DB 540 qrgmflvrqmfalfdmmyssedecrlknwqvldsvrkevaviqppeynrnfansfghl 599  
 QY 626 -AGGYDCQYGYGLWSEVFSMDMFYSCFKKEGIMNPVGMKYRNILKPGSLDGMDLHN 684  
 DB 600 faggysagysyawaevistda-yaafesddv-aatgrfwgeillavdgssraaesfka 657  
 QY 685 FLKREPNOQAFILMSRG 700  
 DB 658 frgepsidallrqsg 673

RESULT 8  
 AAB25647  
 ID AAB25647 standard; Protein; 678 AA.  
 XX AAB25647;  
 AC AAB25647;  
 DT 04-DEC-2000 (first entry)  
 DE N. meningitidis amino acid sequence ml28-1.pap SEQ ID NO:1017.  
 XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW Meningococcus B; MenB.  
 XX Neisseria meningitidis.  
 OS Neisseria meningitidis.  
 XX WO200022430-A2.  
 PN 20-APR-2000.  
 PD 08-OCT-1999; 99WO-US23573.  
 PF 09-OCT-1998; 98US-0103794.  
 PR 30-APR-1999; 99US-0132068.  
 XX (CHIR ) CHIRON CORP.  
 PA Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
 PI Rappuoli R, Pizza M;  
 XX WPI; 2000-318079/27.  
 DR N-PSDB; AAA81287.  
 DR Isolated nucleotide sequences of Neisseria meningitidis which can be  
 DR used in the diagnosis and treatment of N. meningitidis infection and  
 DR other Neisserial infections, for example, N.gonorrhoea -



[illegible]









Query Match	18.9%; Score 694; DB 21; Length 678;
Best Local Similarity	28.2%; Pred. No. 4.4e-51;
Matches 190; Conservative 131; Mismatches 309; Indels 44; Gaps 18;	
QY 55	EQIKTRTEELIVQT--KQVYDVGMLGIEEVT-YENCLOALADVEVKYIVERT-----ML 106
Db 16	dqiktedikpalqataaeareqiaaikaqhtgwantvepltg-----itervgrigvv 70
QY 107	DFPQHVSSEKVEAASTEADKRLSRDIEMSMRGDIPERIVHLOETCDLGKIKPEARRYL 166
Db 71	shlnsvtdtpelraaynelmpeltvftteigqdielynrftkiknspefdtlshaqtkl 130
QY 167	EKSIMGKRNGLHLPQEQVQNEIKSMKRMSELICIDFNKN-LNEDDTFLVF--SKAELGAL 223
Db 131	nhdldrfvlsgaeelppeqgaeklaqltqegaksafsqnvladatdafgyfddaaaplgi 190
QY 224	PDP----FIDSLEKTDQDDKYKITLKYPHYFPVPMKCCIPETRRMEMAFNTRCKE----- 274
Db 191	pedalamfaaaasegktygkigqlqiphylavigyadnrkireqiyrayvtraseisddg 250
QY 275	--ENTILQOLLPLRTKVAKLGLGYSTHADVFLEMTAKSTSRVTAFLDLSQKLPLGEA 332
Db 251	kfdntanidrtlenalqtakllgfknyaelslatkmadtpeqvlfnldlarrakpyaek 310
QY 333	EREFILNLKKCKDRGFEDGKINADWLIYYMTQTEELKYSIDQEFLEKFFPIEVVTEG 392
Db 311	d----laevkafareslgl---adlqpdwldgyageklreaakyafsetevkyfpgvkving 364
QY 393	LLNTYQELLGLSFEQMTDAHVWNKSVTLTVKDKATGCVLGGQFVLDLYPREGYNHAACF 452
Db 365	lfaikkllyigftek-vpvhkdvryfelq--ngetiggyvmdlyaregkrggawm 421
QY 453	GLQPGCLLDGSRMAVAALVNFSPQVAGRPSLLRRHDEVRTYFHEFGVHMQICAQTD 512
Db 422	dygrrfrfsdgtlptaylvcnftppvggkearlshdeiltifhetghlhlqtqde 481
QY 513	ARFSGTN-VETDFEVPQSOMLENWVDVSLRLSKHYKDGSPDIADLLEKLVASRLVNT 571
Db 482	lvgsglmgvewdaveipsgmfewynvlagmsheetgvplpkelfdkmlaaknfr 541
QY 572	GLTLRQIVLSKQDLSLHNTSLDAASEYAKYCSFI---LGVAAATPTGNMPPA-TFGHL-A 626
Db 542	gmflvrgmefaldmmlyseddegriknwgqvlsvrkevavvrpbeynrfansfghifa 601
QY 627	GGYDQGYGYLWSEVFSMDMFYSCPKKEGIMNPEVGMKYNRLILKPGGSLDGMMLHNF 686
Db 602	gysagysyawaevlsada-yaafesddv-aatgkrfwqeilavggssaaesfkaf 659
QY 687	KREPNOKAFILMSRG 700
Db 660	grepsidallrhsg 673
RESULT 15	
AAB25646	
ID AAB25646	standard; Protein; 678 AA.
XX AAB25646;	
AC AAB25646;	
DT 04-DEC-2000	(first entry)
XX N. meningitidis	amino acid sequence a128.pep SEQ ID NO:1015.
DE N. meningitidis	Neisseria gonorrhoeae; genome; immunogenic;
XX antigen; vaccine; diagnosis; infection; antibacterial; identification;	
KW Meningococcus B; MenB.	
XX Neisseria meningitidis.	
OS Neisseria meningitidis.	
XX WO200022430-A2.	
PN	
XX	

20-APR-2000.	
XX 08-OCT-1999;	99WO-US23573.
XX 09-OCT-1998;	98US-0103794.
PR 30-APR-1999;	99US-0132068.
XX (CHIR ) CHIRON CORP.	
XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;	
PI Masignani V, Galcotti C, Mora M, Ratti G, Scarselli M, Scarlato V;	
PI Rappuoli R, Pizza M;	
XX WPI: 2000-318079/27.	
DR N-PSDB; AAB81286.	
XX Isolated nucleotide sequences of Neisseria meningitidis which can be	
PT used in the diagnosis and treatment of N. meningitidis infection and	
PT other Neisserial infections, for example, N.gonorrhoea -	
XX Claim 14; Page 95; 1760pp; English.	
PS The present invention describes methods of obtaining immunogenic	
XX proteins from Neisseria genomic sequences. AAB81453 to AAB82414	
CC represent specifically claimed Neisseria meningitidis genomic DNA	
CC sequences; AAB81260 to AAB81303 and AAB25620 to AAB25663 represent	
CC Neisseria DNA sequences and their corresponding proteins; AAB81254 to	
CC AAB81259 and AAB81304 to AAB81321 represent PCR primers used in the	
CC isolation of Neisseria meningitidis DNA sequences; and AAB81322 to	
CC AAB81452 represent Neisseria meningitidis MenB polynucleotide ORF	
CC sequences, which are all used in the exemplification of the present	
CC invention. The nucleic acid sequences, protein sequences, and antibodies	
CC against them, can be used in the manufacture of a composition. The	
CC composition can be used as a medicament (or in the manufacture of a	
CC medicament) for treating, preventing or diagnosing infection due to	
CC Neisserial bacteria. For example, some of the identified proteins could	
CC be components of vaccines against Meningococcus B; against all serotypes;	
CC and/or against all pathogenic Neisseriae. Identification of sequences	
CC from the bacterium will also facilitate production of biological probes,	
CC particularly organism-specific probes. Attempts to make efficacious	
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.	
CC Multivalent vaccines have also been tried but none have successfully	
CC overcome antigenic variability. The provision of further, complete	
CC sequences may provide an opportunity to identify secreted or surface	
CC exposed proteins that may be presumed targets for the immune system and	
CC which are not antigenically variable or at least more conserved than	
CC other more variable regions.	
XX Sequence 678 AA;	
SQ	

Query Match	18.9%; Score 694; DB 21; Length 678;
Best Local Similarity	28.2%; Pred. No. 4.4e-51;
Matches 190; Conservative 131; Mismatches 309; Indels 44; Gaps 18;	
QY 55	EQIKTRTEELIVQT--KQVYDVGMLGIEEVT-YENCLOALADVEVKYIVERT-----ML 106
Db 16	dqiktedikpalqataaeareqiaaikaqhtgwantvepltg-----itervgrigvv 70
QY 107	DFPQHVSSEKVEAASTEADKRLSRDIEMSMRGDIPERIVHLOETCDLGKIKPEARRYL 166
Db 71	shlnsvtdtpelraaynelmpeltvftteigqdielynrftkiknspefdtlshaqtkl 130
QY 167	EKSIMGKRNGLHLPQEQVQNEIKSMKRMSELICIDFNKN-LNEDDTFLVF--SKAELGAL 223
Db 131	nhdldrfvlsgaeelppeqgaeklaqltqegaksafsqnvladatdafgyfddaaaplgi 190
QY 224	PDP----FIDSLEKTDQDDKYKITLKYPHYFPVPMKCCIPETRRMEMAFNTRCKE----- 274
Db 191	pedalamfaaaasegktygkigqlqiphylavigyadnrkireqiyrayvtraseisddg 250
QY 275	--ENTILQOLLPLRTKVAKLGLGYSTHADVFLEMTAKSTSRVTAFLDLSQKLPLGEA 332
Db 251	kfdntanidrtlenalqtakllgfknyaelslatkmadtpeqvlfnldlarrakpyaek 310
QY 333	EREFILNLKKCKDRGFEDGKINADWLIYYMTQTEELKYSIDQEFLEKFFPIEVVTEG 392
Db 311	d----laevkafareslgl---adlqpdwldgyageklreaakyafsetevkyfpgvkving 364
QY 393	LLNTYQELLGLSFEQMTDAHVWNKSVTLTVKDKATGCVLGGQFVLDLYPREGYNHAACF 452
Db 365	lfaikkllyigftek-vpvhkdvryfelq--ngetiggyvmdlyaregkrggawm 421
QY 453	GLQPGCLLDGSRMAVAALVNFSPQVAGRPSLLRRHDEVRTYFHEFGVHMQICAQTD 512
Db 422	dygrrfrfsdgtlptaylvcnftppvggkearlshdeiltifhetghlhlqtqde 481
QY 513	ARFSGTN-VETDFEVPQSOMLENWVDVSLRLSKHYKDGSPDIADLLEKLVASRLVNT 571
Db 482	lvgsglmgvewdaveipsgmfewynvlagmsheetgvplpkelfdkmlaaknfr 541
QY 572	GLTLRQIVLSKQDLSLHNTSLDAASEYAKYCSFI---LGVAAATPTGNMPPA-TFGHL-A 626
Db 542	gmflvrgmefaldmmlyseddegriknwgqvlsvrkevavvrpbeynrfansfghifa 601
QY 627	GGYDQGYGYLWSEVFSMDMFYSCPKKEGIMNPEVGMKYNRLILKPGGSLDGMMLHNF 686
Db 602	gysagysyawaevlsada-yaafesddv-aatgkrfwqeilavggssaaesfkaf 659
QY 687	KREPNOKAFILMSRG 700
Db 660	grepsidallrhsg 673

Db 251 kfdntanidrtlenalqtakllgfknyaelslatkmdtpeqvlnflhdlarrakpyaek 310  
QY 333 BREFILNLKKCKEKGFEYDCKINAWDLYYMYTQTEELKYSIDOEFLKEYEPIEVVTEG 392  
Db 311 d---laevkafareslgl---adlcpwdlgyageklreakyafsecevkkyfpvgkvlnq 364  
QY 393 LNTYQELLGLSFEQMTDAHVWNKSVTLYTVKDKATGEVLGQFYLDLYPREGKYNHAACF 452  
Db 365 lfaqlkkllygigftek-vpwwhkdvryfelq--ngetiggyvmdlyaregkrggawmn 421  
QY 453 GLQPGCLLPDGSRRMAVALVNFQSPVAGRPSLLRHDDEVRTYFHEFGHVHMHOICAQTDf 512  
Db 422 dykgrrrfsdgtlqiptaylvcnfppvggkearlshdeilltfhetghlhhlltqvde 481  
QY 513 ARFSGTN-VETDFVEVPQOMLENWVWDLSRLSKHYKDGSPiADDDLLEKLVASRLVNT 571  
Db 482 lqvsqingvewdaveipsgfmenfweynvlaqmsaheetgvplpkelfdkmlaaknqfr 541  
QY 572 GLTLRLQIVLSKVDQSLHTNTSLDAASEYAKYCSEI---LGVAATPGTNMPA-TFGHL-A 626  
Db 542 gmflvrqmfalfdmmyiseddegrlknwqqvldsvrkevavvrppeynrfansfghifa 601  
QY 627 GGYDQYGYGLWSEVFSMDMFYSCFKKGIMNPEVGMKYRNLIILKPGSLDGMMLHNFL 686  
Db 602 ggyysagyysawaevlsada-yaafeesddv-aatgkrfwqellavggssraaesfkafr 659  
QY 687 KREPNOKAFLMSRG 700  
Db 660 grepsidalrrhsg 673

Search completed: January 14, 2002, 18:55:59  
Job time: 1043 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2002, 19:00:21 ; Search time 91.83 Seconds  
(without alignments)  
5216.171 Million cell updates/sec

Title: US-09-833-782-1

Perfect score: 2115

Sequence: 1 atgatgcgcgggtgcctttt.....gagcgctgcgtcgtcgtga 2115

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*

3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*

4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*

5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1590.8	75.2	2148	5	PCT-US92-00731-12
2	790.8	37.4	2070	1	US-07-766-351-6
3	790.8	37.4	2070	1	US-08-059-032-6
4	790.8	37.4	2070	5	PCT-US91-07290-6
5	86.8	4.1	723	4	US-08-998-416-839
6	61	2.9	7218	1	US-08-232-463-14
7	42	2.0	75	1	US-07-766-351-13
8	42	2.0	75	1	US-08-059-032-13
9	42	2.0	75	5	PCT-US91-07290-13
10	41.4	2.0	164	1	US-07-766-351-16
11	41.4	2.0	164	1	US-08-059-032-16
12	41.4	2.0	164	5	PCT-US91-07290-16
13	36.2	1.7	1821	4	US-09-056-105-23
14	36.2	1.7	1896	1	US-08-253-503-1
15	36.2	1.7	1896	1	US-08-796-883-1
16	36.2	1.7	1896	2	US-08-611-273B-1
17	36.2	1.7	1896	2	US-08-531-864-1
18	36.2	1.7	1896	2	US-08-373-636C-1
19	36.2	1.7	1896	3	US-08-602-506A-1
20	36.2	1.7	1896	4	US-09-266-294-1
21	36.2	1.7	1896	4	US-09-179-281-1
22	36.2	1.7	1896	4	US-09-056-105-25
23	35.4	1.7	3095	6	5231168-1
24	35.2	1.7	5741	1	US-07-706-699-4
25	35.2	1.7	5741	1	US-07-998-931-4
26	34.4	1.6	651	2	US-08-929-418-1
27	34.2	1.6	5049	1	US-08-336-345-1

28 34.2 1.6 5049 1 US-08-336-345-2 Sequence 2, Appli  
29 34.2 1.6 5049 2 US-08-647-655-1 Sequence 1, Appli  
30 34.2 1.6 5049 2 US-08-647-655-2 Sequence 2, Appli  
31 34 1.6 7125 1 US-07-745-206A-1 Sequence 1, Appli  
32 34 1.6 7125 2 US-08-311-363-1 Sequence 1, Appli  
33 34 1.6 7635 1 US-08-455-543A-1 Sequence 23, Appli  
34 34 1.6 7635 2 US-08-193-078B-1 Sequence 1, Appli  
35 34 1.6 7635 2 US-08-223-305C-1 Sequence 29, Appli  
36 34 1.6 7635 2 US-08-223-305C-1 Sequence 1, Appli  
37 34 1.6 7635 2 US-08-223-305C-23 Sequence 23, Appli  
38 34 1.6 7635 2 US-08-149-097D-1 Sequence 1, Appli  
39 34 1.6 7635 3 US-08-949-386-1 Sequence 1, Appli  
40 34 1.6 7635 3 US-08-450-562-1 Sequence 1, Appli  
41 34 1.6 7486 3 US-08-475-886-5 Sequence 5, Appli  
42 33.8 1.6 7486 4 US-08-397-232-3 Sequence 3, Appli  
43 33.8 1.6 7486 3 US-08-475-886-3 Sequence 3, Appli  
44 33.8 1.6 7486 3 US-08-475-886-1 Sequence 1, Appli  
45 33.8 1.6 7486 3 US-08-475-886-1 Sequence 1, Appli

## ALIGNMENTS

RESULT 1  
PCT-US92-00731-12  
; Sequence 12, Application PC/TUS9200731  
; GENERAL INFORMATION:  
; APPLICANT: Kawabata, Shunichiro  
; APPLICANT: Davie, Earl W.  
; TITLE OF INVENTION: MICROSOMAL ENDOPEPTIDASE  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed and Berry  
; STREET: 6300 Columbia Center  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: US  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/00731  
; FILING DATE: 19920128  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/646,997  
; FILING DATE: 28-JAN-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 990008.415PC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206)622-4900  
; TELEFAX: (206)682-6031  
; TELEX: 3723836  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2148 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FRAGMENT TYPE: N-terminal  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 11..2075  
; PCT-US92-00731-12

Query Match 75.2%; Score 1590.8; DB 5; Length 2148;

	Best Local Similarity	87.2%;	Pred. No. 0;		Matches	1728;	Conservative	1;	Mismatches	247;	Indels	6;	Gaps	1;
Qy	1	atgatccgcgggtgcttttggctgtgcgaagcctccgcagagtggtggttcaggatt	60											
Db	167	ATNATCCCGCGTGCTTTTGGCTGTGNGAGGCCCTCACAGGGTGTGTGTCACGATT	226											
Qy	61	ttaetcaagaatgacgttagaagaagaagtgtctctctcttcaggcaaatgcttcctat	120											
Db	227	CTATTCAAATGACATTAGGAAGAAGATGATGCACCTCTTCAGCAGATTCTTCCTAT	286											
Qy	121	actgtggtcgccgaataatgtttaagatggatctttaccacagagcaaaataaacaca	180											
Db	287	ACTGCAGCTGGCAGGAATGTTTTAAGATGGATCTTTCACCAGAGCAAAATCAAGACA	346											
Qy	181	actgagagctcatgtgcagaccaaacagggtfacgattgctgttggaatcctcggtatt	240											
Db	347	ACTGAGGAGCTCATPTGGCGAGACCAGCAGGTGTATGTTCTGTGGAAATGCTGGATC	406											
Qy	241	gaggaagtaacctacgagaactgtctgcagcacctgccagatgtagaagtaaagtata	300											
Db	407	AAGGACGTGACTPACGAGAACTGTTGCAGCGCTGGCAGATGTGGAAGTGAATACATA	466											
Qy	301	gtggaagggaccatgttagactttccccagcatgtactctgcacaaagaagtacgagca	360											
Db	467	GTGAAAAGAACCATGTGTAGACTTTTCTCCAGCATGTTCTACTGCAGACAAGTACGGSCA	526											
Qy	361	gcaagtacagaagcagacacaaaagactttcttgttatattgagatgagcatgagagga	420											
Db	527	GCAAGTACAGAAGCGGACACAGAGCGTTTCTCGCTTTGTATTTGATGTAGATGAGCATGAGAGAA	586											
Qy	421	gatatatgaagaattgttcatttacagaaaacctgtgctcgtgggaagataaaacct	480											
Db	587	GATATATTTACAGAAATCGTTTCACTTACAGAAACCTGTGATCTGGAGAAATTAAGCCG	646											
Qy	481	gagccagacgatacttggaaaagtccaattaaatgggaaaaagaaatgggtccatctt	540											
Db	647	GAECGAGACGACTACTTGGAAAAGTCAGTTTAAATGGGAAGSNATGGCTCCACCTT	706											
Qy	541	cctgaacaagtacagaatgaatcaaatcaatgaagaaaagaatgagtgcctatgtatt	600											
Db	707	CCNNNAGAAGTACAGAATGANATCAATCAATGAAGAAAGAATGAGTGCAGCTGTGATT	766											
Qy	601	gatttaacaaaaacctcaaagagatgataccttcctgtattttccaaagctgaacct	660											
Db	767	GACTTTTAAAAAACCTCAATNAGATGATACCNFTCCNTGTGTTTTCCAGGCTGAGCTT	826											
Qy	661	gggtcctctcgtgatgttccattgacagtttagaaaaagacagatgatacaagtataaa	720											
Db	827	GGTCTCTTCTCGATGATTCATTGACAGTTTAGAAAAGATGGATGATGACAAAGTATAA	886											
Qy	721	ataccttaaaatatccacacattttccctgcatggaagaaatgttgtatcccctgaaccc	780											
Db	887	ATTACCTTAAAAATATCCACATATTTTCTGTCAATGAAGAAATGTTGTTCTCTGAAACC	946											
Qy	781	agaagaaggatggaatgcttttaatacagaagtgcaagaggaacacacataattttg	840											
Db	947	AGAAGGAGATGGAATGGCTTTTAATACAGATGCAAGAGGNNAACACCGTCATTCG	1006											
Qy	841	cagcagctactcccactgcgaaccaagtgcccaaaactactcgtttatagcacacatgct	900											
Db	1007	CAGCAGCTACTCCCACTGCGGGCCCCAAGTGCCCAAACTCCTGGGCTATAGCACACATGCT	1066											
Qy	901	gacttcgtcccttgaatgaacactgcgaagagacaaagccgcgttaacagcytttcttagat	960											

APPLICANT: Seubert, Peter A.  
 APPLICANT: Dovey, Harry F.  
 APPLICANT: McConlogue, Lisa C.  
 APPLICANT: Little, Sheila P.  
 APPLICANT: Johnstone, Edward M.  
 TITLE OF INVENTION: Amyloidin Protease and Uses Thereof  
 NUMBER OF SEQUENCES: 18  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Athena Neurosciences, Inc.  
 STREET: 800F Gateway Blvd.  
 CITY: South San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94080  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/766.351

	Query Match	37.4%	Score 790.8;	DB 1;	Length 2070;
	Best Local Similarity	62.4%;	Pred. No. 7.3e-217;		
	Matches 1236;	Conservative 1;	Mismatches 743;	Indels 0;	Gaps
Qy	128	ctggcagaatagttttaagatggatctttcacccagacaaaattaaacaagaactgaag	187		
Db	56	CTGTGTTAAACGACCTCGGTGGACCTGTGAGTCCGACAGCATAGAGGACGCGACCAGGG	115		
Qy	188	agtcattgtgcagacaaacagdgtagcatgtcttggaatgctcgggtattgaggaag	247		
Db	116	AGCTCATCGAGCAGACCAAGCGGTGTATGCCAGGTTGGCACCCAGGAGTTTGAGGACG	175		
Qy	248	taacttacgagaactgtctgcagcactgacagatgtagaagtaaaagtatatagtggaaa	307		
Db	176	TGTCCTACGAGAGCACCGCTCAAGCGCTGCGCGATGTGGAGGTCACCTACACAGTTTCAG	235		
Qy	308	ggaccatgctagactttcccagcatgtatcctctgcacaaagaagtacgagcagcaagta	367		
Db	236	GGATATATCTTGACTTTCCTCCACAGCATGTTTCCCTCCCAAGGACATCCGGACAGCCAGCA	295		
Qy	368	cagaagcagacaaaagactttctcgtttgatatgagatgagcatgagagagatatat	427		
Db	296	CAGAGGCCGACAAGAAGCTCTCTAGTTCGACCTGGAGATGAGCATGAGGGAGAGCGTGT	355		
Qy	428	ttagagaaattgttcatttacagaaaacctgtgactcggggaagataaaacctgaggcca	487		
Db	356	ACCAGAGGATCGTGTGGCTCCAGGAGAAAGTTTCAGAAGGACTCACTGAGGCCCGAGCGCTG	415		
Qy	488	gacatacttggaaaaagtcattataaatggggaagaagaatgggctccatcttccctgaac	547		
Db	416	CGCGGTACCTGGAGCGGCTAATCAAGCTGGCGCGGAGAATGGCTTCACCTCCCCAGAG	475		



QY 1028 aggaatgcaaacacaggggttttgaatgatgaggaatcaatgcctggggtatatt 1087  
Db 956 CGGAGTGCAGCGCGCGCTGCGCTTCGACGCGCGATCGCTGCGGACATCGGCT 1015  
QY 1088 actacatgactcagacagaggaactcaagttatccatgaccagaagttcctcaaggaa 1147  
Db 1016 ACTATGATCAACCAAGTGGAGGAGACGCGCTACTGGGTGACCAAGCACTGCTCAAGGAGT 1075  
QY 1148 acttcccattgaggtggtcactgaaggttctgtgaaacacacctaccagaggttctgggac 1207  
Db 1076 ACTTCCCGTGCAGTGGTTCACGACGGCTGCTGGGATCTACCAAGAGCTCCTGGGGC 1135  
QY 1208 ttctattgaaacaatgacagatgctatgttttggaaacaagagtgttacactttatctg 1267  
Db 1136 TGGCCTTCCACCACGAGGAGGCGCGCTGCTGGCATGAGACGCTGGCGCTTACACG 1195  
QY 1268 tgaagataaactcagagagaagtattggacagttctatttgaacctctatccaagg 1327  
Db 1196 CGAGGACGCGCGCTCGGGGAGGTGGTGGCAAGTCTACCTGGACCTGTACCCGCGG 1255  
QY 1328 aaggaaaatacaatcatcgcgctctgtctcagctgctgctctctgctgatg 1387  
Db 1256 AGGAAAGTACGGGACGCGGCTGCTTGGCTGACGCGCGCTGCTGGCGGAGGATG 1315  
QY 1388 gaagccggaatgagtgagtggtgctgctgctggtgaaacttctcacagccagtggcaggtc 1447  
Db 1316 GGAGCGCGCAGATCGCATCGCGGCGATGTCGCCAATTCACCAAGCCACAGCGGAGC 1375  
QY 1448 gtccctctctctgagacacgacgaggtgaggaacttaactttatgagtttgctcagtg 1507  
Db 1376 CCGCTCGCTGCTGACGATGACGAGGTGGAGACCTACTTCATGAGTTTGGCCACGTA 1435  
QY 1508 tgcatacagattgtcacagactgatttgcacgactttagcaggaacaagtgtgaaactg 1567  
Db 1436 TGCACAGCTCTGCTCCAGCGCGAGTTCGCCAATTCACGCGGACCCACAGTGGAGCGG 1495  
QY 1568 actttgtagagtgccatgcgaatgcttgaataatgggtgagcgtcgattccctcc 1627  
Db 1496 ACTTTGTGGAGCGCGCTGCGAGATGCTGGAGAACTGGGTGGGAGGAGCGCGCTGC 1555  
QY 1628 gaagattgtcaaaaacattataaagatgaagccctattgagacgatctgttgaaaaac 1687  
Db 1556 TCGCGATGTCGCGGCACTACCGACAGGACGCGCGCTGCCCGGAGCTCCTGGAGAAAG 1615  
QY 1688 ttgttgccttagctgtcaacagacttcttcaacctgcgccagattgttttgagca 1747  
Db 1616 TCATTGAGTCCCGGAGGCGCAACACAGGCGCTTTCAGCGCTGCGCCAGATCGCTCGCCA 1675  
QY 1748 aagttgatcagtcctctcatcaacaacacatcgctggtgctgcaagtgaaatgcaaat 1807  
Db 1676 AGGTGGACCGGCGCTGACACGCGAGCGGAGCGAGACCCCGCCAGAGATGCGCGGC 1735  
QY 1808 actgctcagaaaattagagttgagctacttccaggcaaaaatgacagctactttg 1867  
Db 1736 TCTGCCAGGAGATCTCGGGTCCCGGCCACGCCAGGAACCAATGCTGCAACCTCG 1795  
QY 1868 gacatttggcaggggatacagatgcccataattatgatatatttggagtgaaagtatttt 1927  
Db 1796 GCGATCTGCGAGGTGGCTACGACGCGCACTAGTACGGGTACCTGTCGAGCGAGGTGATT 1855  
QY 1928 ccattgatattgtttacagctgttttaaaaagaagggaataatgaatccagaggtgaa 1987  
Db 1856 CCATGGACATGTTCCACACGCGCTTCACGAGGAGGGTGTCTCTTAACAGCAAGGTTGCA 1915  
QY 1988 tgaatacagaaaacctaactgaaacctgggggactctctgagcggaatggacatgctcc 2047  
Db 1916 TGGATTACAGAGCTGCATCTCTGACCGCGGCTTCGAGGATGCCAGCGCATGCTGA 1975  
QY 2048 acaatttctgaaacgtagacccaacaaaagcgttctaatgagtagagcctgcatg 2107  
Db 1976 GCGGCTTCTTGGGCGCTGACCCCAAGCAGGAGCGCTTCTCTCTGAGCAAGGGGCTGCAAG 2035

RESULT 4  
PCT-US91-07290-6  
; Sequence 6, Application PC/TUS9107290  
; GENERAL INFORMATION:  
; APPLICANT: Sinha, Sukanto  
; APPLICANT: Seubert, Peter A.  
; APPLICANT: Dovey, Harry F.  
; APPLICANT: McConlogue, Lisa C.  
; APPLICANT: Little, Sheila P.  
; APPLICANT: Johnstone, Edward M.  
; TITLE OF INVENTION: Amyloidin Protease and Uses Thereof  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Athena Neurosciences, Inc.  
; STREET: 800F Gateway Blvd.  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US91/07290  
; FILING DATE: 19911004  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy, Lisabeth Felix  
; REGISTRATION NUMBER: 31547  
; REFERENCE/DOCKET NUMBER: 17796-002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 877-0900  
; TELEFAX: (415) 877-8370  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2070 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
PCT-US91-07290-6

Query Match 37.4%; Score 790.8; DB 5; Length 2070;  
Best Local Similarity 62.4%; Pred. No. 7.3e-217;  
Matches 1236; Conservative 1; Mismatches 743; Indels 0; Gaps 0;

QY 128 ctggcagaaatgtttaagatggggtatctttccaccagagcaaatataaagaactgag 187  
Db 56 CTGTGTTAAACACCTCGGTGGGACCTGAGTGCCACAGATAGAGGAGCGCACGAG 115  
QY 188 agctcattgtgcagaccacaaacaggtgtgacgatgctgttggaatgctcgtgattgagga 247  
Db 116 AGCTCATCGACGACCAAGCGCGTGTATGACCAAGTTGGCACCAGAGTTTGAGGAC 175  
QY 248 taacttacgaaactgtctgcaggcactggcagatgtagaagataagatatagtggaaa 307  
Db 176 TGTCTACGACGACGCTCAAGCGCTGCGCGATGTGGAGTCACTACACAGTTTCA 235  
QY 308 ggaccatgtagactttcccccagcatgtatctctgacaaagaagtacgacgagcaagta 367  
Db 236 GGAATATCTTGAATCTCCCGACGATGTTTCCCGCTCCAAAGGACATCCGACACGCA 295  
QY 368 cagaagcagacaaagaactttctcgttttgatattgagatgagcagagagatatat 427  
Db 296 CAGAGCGCGCAAGAAGCTCTCTGAGTTTCGAGCTGGAGATGAGATGAGGAGGAGGTGT 355  
QY 428 ttgagagaattgtttcatttacaggaaacctgtgatctgggggaagataaaacctgagcca 487

Db 356 ACCAGGATCGTGTGGCTCCAGGAGAAAGTTCAAGAGACTCACTAGGCGCCGAGGCTG 415  
Qy 488 gacgatacttgaaaaagtaattaaaaatgggaaaaaagaatgggctccatcttccctgaac 547  
Db 416 GCGGTACTCTGGAGGGCTAATCAAGCTGGCGGAGAAATGGGCTTCACCTCCCCAGAG 475  
Qy 548 aaglacagaatgaatacaatacaatgaagaaaaaagaatgagtgatgtattgatttta 607  
Db 476 AGACTCAGGAAACATCAACGCATCAGAGAGAGAGCTGAGCCTTCTGTGCATCGACTCA 535  
Qy 608 acaaaaacctcaatgagatgataaccttccctgttattttccaaggtgaaacttggtgctc 667  
Db 536 ACAAGAACTGAAGGAGGACACGACCTTCTTCCCTCCCTCACGCTCCAGGAGTAGGAGGC 595  
Qy 668 ttctgtagatttcattgacaggtttagaagaagacagatgatacaagttataaaattacct 727  
Db 596 TCCCGGAGGACTTCTGNACTTCCCTGGAGAAGATGGAGGACGGAAGTTGAAGGTCAACC 655  
Qy 728 taataatccacactatttccctgtcatgaagaaatgttgtatccctgaaaccagaagaa 787  
Db 656 TCAAGTACCCCCATTACTTCCCTCCCTCTGAAAGAAATGCCACGTGCTGAGACAGGAGGA 715  
Qy 788 99atggaataatggcttttaatacaaggtgcaagagggaaaaaacacacataattttgcagcgc 847  
Db 716 AAGTGGAGGAGGCTTCAACTGCGCGGTGCAAGGAGGAGAACTGGCGCTATCCTCAAGGAGC 775  
Qy 848 taetcccaactgcgaacaaagtggccaaactactcgtgttatagcacacatgtaacttcg 907  
Db 776 TGGTGACCTCGCGGGCCCAAGTCCCGCTGCTGGGGTTCCACACGACGCCGACTATG 835  
Qy 908 tccttgaataaacaactgaaagagcaaacgcgcgttaacagcyctttctagatgatttaa 967  
Db 836 TCCTGGAGATGAACATGSCCAAGACCAAGACCGACCGTGGCCACCTTCTAGATGAGCTGG 895  
Qy 968 gccagaagttaaacaccttggttggaagcagagacagaggtttatttgaatttggaagaaa 1027  
Db 896 CGCAGAACTGAAGCCCTCGGGGGAGCAGGAGCGTGGCGTGTGATCTGGAGCTGAAGCGTG 955  
Qy 1028 aggaatgcaagacaggggttttgaatatgatggaaaaatcaatgcctggatctatatatt 1087  
Db 956 CGGAGTGGACGCGCGGGGCTGCGCTTCGACGCGCGCATCCGTGCTTGGGACATYGGCT 1015  
Qy 1088 actacatgactcagacagaggaactcaagtaattccatagaccacagagtctcacaagaaat 1147  
Db 1016 ACTACATGAACCAAGTGGAGGAGACGCGCTACTGCTGGACCAACACTGCTCAAGGAGT 1075  
Qy 1148 acttcccaatgaggttggttcaactgaaggttgctgaaacacacacacagaggttggtgac 1207  
Db 1076 ACTTCCCGCTGACAGTGGTTCACGACGCGGCTGCTGGGCATCTACACGAGGCTCCTGGGGC 1135  
Qy 1208 ttcatattgaacaaatgacagatgctcatgttttgaaacaagagtgttacacttttactg 1267  
Db 1136 TGGCTTCCACCAAGAGGAGGCGCCAGTGGCTGGCATGAGAGCTGGCGGCTCTACACCG 1195  
Qy 1268 tgaaggataaagctacagggagaagatttgggacaggtttctatttggactctatccaaggg 1327  
Db 1196 CGAGGAGCGCGCTCGGGGAGGTGGTGGCAAGTCTTACTTGGACCTGATCCCGCGGG 1255  
Qy 1328 aaggaaaaatacaatcatgcggcctgcttccgtctccagcctggtgcttctgctgatg 1387  
Db 1256 AAGGAAAGTACGGGACGCGGCTTGGCTTGGCCCTGCAAGCCGCGCTGCTGGGAGGATG 1315  
Qy 1388 gaagccggatgagcagtgaggtgaggaacttactttcatgagtttggttcacgtga 1447  
Db 1316 GGAGCCGCAGATCGCATCGCGCCATGGTGGCCCACTTCCCAAGCCCAACAGCCGACG 1375  
Qy 1448 gtccctctctctgagacagaacagaggtgaggaacttactttcatgagtttggttcacgtga 1507  
Db 1376 CGCCTCGCTGCTGCAGCATGACGAGGTGGAGACCTACTTCCATGAGTTGGCCACGTGA 1435  
Qy 1508 tgcatacgatttgcacagactgattttgcacgatttagcgggaaaaatgtggaaactg 1567  
Db 1436 TGCACAGCTCTGCTCCCGCGGAGTTCCGCAATGTTCAAGCGGGACCACTGTGGAGCGGG 1495

Qy 1568 actttgtagagtgccatcgcaaatgcttgaataatgggtgtggagctgcattccctcc 1627  
Db 1496 ACTTTGTGGAGGCGCCCTGCGAGATGCTGGAGAACTGGGTGTGGAGCAGAGCGGTGC 1555  
Qy 1628 gaagattgtcaaaacattataaagatgaagccctatttgacagacgactgtcttgaaaaac 1687  
Db 1556 TCGGATGTGCGGCACTTACCGCACAGCAGCGCGGTGCCCCGGGAGCTCCTTGAGAAAGC 1615  
Qy 1688 ttgtgtcttagagctggtcaaacacaggtcttctgacctgctgcacagattgttttgagca 1747  
Db 1616 TCATTGAGTCCCGCAGCGCAACACAGGCGCTCTTTCAGCTCGCCAGATCCTCTCGCCA 1675  
Qy 1748 aagttgatcagctctcttcataccaacacatcgctggtgctgcaagtgaatatgccaaat 1807  
Db 1676 AGGTGGACCAAGCCCTGACACAGCAGCGGACGAGACCCCGCCGAGGAGTATGCGCGGC 1735  
Qy 1808 actgctcagaataattaggagttgcagctactcactccaggcacaataatgccagctaccttg 1867  
Db 1736 TCTCCAGGAGATCCTCGGGGTCCCGGCCACGCGCAGGAACCAACATGCTGCAACCTTCG 1795  
Qy 1868 gacatttggcaggggatacagatggccaataattatgatatcttcttgagtgagtgaaatttt 1927  
Db 1796 GCCATCTGGCAGGTGGCTACGACGCCCACTACTACGGGTACTCTGGAGCGAGGTGATT 1855  
Qy 1928 ccattgatatattttcacagctgttttaaaaaagagggaataatgaatccagaggttgaa 1987  
Db 1856 CCATGGACATGTTCCACACGCGCTTCAAGCAGGAGGTTGCTCTGAACACGCAAGGTTGCA 1915  
Qy 1988 tgaatacagaaaactaatcctgaacctggggtatctctgagcgcatggacatgctcc 2047  
Db 1916 TGGATTACAGAAGTGCATCTGAGACCGCGGTTCGAGGATGCCAGCGCATGCTGA 1975  
Qy 2048 acaatttctgaaacgtgagccaaacaaagaggttctcttaataagtagagagcctgatg 2107  
Db 1976 GCGCTTCTTGGCGGTGACGCCAAGCAGGACGCTTCTCTCTGAGCAAGGGGTGCAGG 2035

## RESULT 5

US-08-998-416-839

: Sequence 839, Application US/08998416

: Patent No. 6239264

: GENERAL INFORMATION:

: APPLICANT: Philippsen, Peter

: APPLICANT: Pohlmann, Rainer

: APPLICANT: Steiner, Sabine

: APPLICANT: Mohr, Christine

: APPLICANT: Wendland, Jergen

: APPLICANT: Knechtle, Philipp

: APPLICANT: Reibschung, Corinne

: TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPPI

: TITLE OF INVENTION: AND USES THEREOF

: NUMBER OF SEQUENCES: 1152

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: No. 6239264artis Corporation

: STREET: 3054 Cornwallis Road

: CITY: Research Triangle Park

: STATE: No. 6239264th Carolina

: COUNTRY: USA

: ZIP: 27709

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: PatentIn Release #1.0, Version #1.30

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/998,416

: FILING DATE: 24-DEC-1997

: CLASSIFICATION: 435

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: CH 0016/97

: FILING DATE: 31-DEC-1996

: ATTORNEY/AGENT INFORMATION:





Db 1085 RRRRRRRRRRRRRRRATCGCAAGCTCCCTCGACCTGCAGCAA 1039

```

RESULT 7 |
US-07-766-351-13
; Sequence 13, Application US/07766351
; Patent No. 5292652
;
; GENERAL INFORMATION:
;
; APPLICANT: Sinha, Sukanto
; APPLICANT: Seubert, Peter A.
; APPLICANT: Dovey, Harry F.
; APPLICANT: McConlogue, Lisa C.
; APPLICANT: Little, Sheila P.
; APPLICANT: Johnstone, Edward M.
;
; TITLE OF INVENTION: Amyloidin Protease and Uses Thereof
;
; NUMBER OF SEQUENCES: 18
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Athena Neurosciences, Inc.
; STREET: 800F Gateway Blvd.
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
;

```

```
Query Match 2.0%; Score 42; DB 1; Length 75;
Best Local Similarity 73.0%; Pred. No. 0.0055;
Matches 54; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
```

Qy 1141 aaggaatacttcccaattgagggtggtcactgaagccttgctgaacacctaccaggagttg 1200  
||||| ||||||| | ||||||| || ||||||| || ||||||| |  
Db 1 AAGGAGTACTTCCTGTGCAGGTGGTCAAGCACGGCTGCTGGGCATCTACAGGAGGTC 60

Qy 1201 ttgggactttcatt 1214  
Db 61 CTGGGCCCTGGCCCTT 74

RESULT 8  
US-08-059-032-13  
: Sequence 13, Application US/08059032  
: Patent No. 5424305  
: GENERAL INFORMATION:  
: APPLICANT: Sinha, Sukanto  
: APPLICANT: Seubert, Peter A.  
: APPLICANT: Dovey, Harry F.

APPLICANT: McConlogue, Lisa C.  
 APPLICANT: Little, Sheila P.  
 APPLICANT: Johnstone, Edward M.  
 TITLE OF INVENTION: Anyloidin Protease and Uses Thereof  
 NUMBER OF SEQUENCES: 18  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Townsend and Townsend Kourie and Crew  
 STREET: One Market Plaza, Steuart Street Tower,  
 STREET: Suite 2000  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94105  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/059,032  
 FILING DATE: 19930507  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Smith, William E.  
 REGISTRATION NUMBER: 30,223  
 REFERENCE/DOCKET NUMBER: 15270-10  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 326-2400  
 TELEFAX: (415) 326-2422  
 INFORMATION FOR SEQ ID NO: 13:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 75 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: CDNA  
 HYPOTHETICAL: YES  
 ANTI-SENSE: NO  
 US-08-059-032-13

Query Match 2.0%; Score 42; DB 1; Length 75;  
Best Local Similarity 73.0%;  
Pred. No. 0.0055;  
Matches 54; Conservative 0; Mismatches 20; Indels

Qy	1141	aaggaataacttcccaattgagggtggtcactgaagccttgc	gaacacctaccaggaggttg	1200
Dh	1	AGGAGTACTTCCCTGTCGAGGTGCTCACACACGGCTGCTGGG	CTATCTACCAAGGAGTTC	60

Qy	1201	ttgggactttcatt	1214
Db	61	CTGGGGCCTGGGCTT	74

RESULT 9  
PCT-US91-07290-13  
Sequence 13, Application PC/TUS9107290  
GENERAL INFORMATION:  
APPLICANT: Sinha, Sukanto  
APPLICANT: Seubert, Peter A.  
APPLICANT: Dovey, Harry F.  
APPLICANT: McConlogue, Lisa C.  
APPLICANT: Little, Shella P.  
APPLICANT: Johnstone, Edward M.  
TITLE OF INVENTION: Amyloidin Prot  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Athena Neurosciences,  
STREET: 800F Gateway Blvd.  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080

```
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US91/07290
/ FILING DATE: 19911004
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Murphy, Lisabeth Feix
/ REGISTRATION NUMBER: 31547
/ REFERENCE/DOCKET NUMBER: 17796-002
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 877-0900
/ TELEFAX: (415) 877-8370
/ INFORMATION FOR SEQ ID NO: 13:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 75 base pairs
/ TYPE: NUCLEIC ACID
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cdna
/ HYPOTHETICAL: YES
/ ANTI-SENSE: NO
/ PCT-US91-07290-13

Query Match 2.0%; Score 42; DB 5; Length 75;
Best Local Similarity 73.0%; Pred. No. 0.0055;
Matches 54; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 1141 aaggaatactcccaattgagtggtgactgaagcttctgctgaacacctaccaggagttg 1200
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 AAGAGTACTTCCTGTGTCAGGTGGTCACGCCAGGCTGCTGGGCATCTACCAAGAGCTC 60

QY 1201 ttgggacttcatt 1214
      ||||| |||||
Db 61 CTGGGCTGGCCTT 74

RESULT 10
US-07-766-351-16
; Sequence 16, Application US/07766351
; Patent No. 5292652
; GENERAL INFORMATION:
; APPLICANT: Sinha, Sukanto
; APPLICANT: Seubert, Peter A.
; APPLICANT: Dovey, Harry F.
; APPLICANT: McConlogue, Lisa C.
; APPLICANT: Little, Sheila P.
; APPLICANT: Johnstone, Edward M.
; TITLE OF INVENTION: Amyloidin Protease and Uses Thereof
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Athena Neurosciences, Inc.
; STREET: 800F Gateway Blvd.
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/766,351
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy, Lisabeth Feix
; REGISTRATION NUMBER: 31547

/
/ REFERENCE/DOCKET NUMBER: 17796-002
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 877-0900
/ TELEFAX: (415) 877-8370
/ INFORMATION FOR SEQ ID NO: 16:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 164 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cdna
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ US-07-766-351-16

Query Match 2.0%; Score 41.4; DB 1; Length 164;
Best Local Similarity 56.1%; Pred. No. 0.012;
Matches 78; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 775 gaaaccagaagaagatggccttttaataacaaggtgcaagaggaacacaccata 834
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 26 GGAAGTGGAGGAGCGCTTCAACTGCGGTGCAAGCGGTGCAAGGAGGAGAACTGCGCT 85

QY 835 attttgcagcagctactccactcgaaccaaggtgccaactactcgtttatagcaca 894
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86 ATCCTCAAGGAGCTGCTGACGCTGCGGGCCAGAAAGTCCGCGCTGCTGGGGTTCCACACG 145

QY 895 catgctgacttccttg 913
      ||||| ||||| |||||
Db 146 CACGCCGACTATGCTCTGG 164

RESULT 11
US-08-059-032-16
; Sequence 16, Application US/08059032
; Patent No. 5424205
; GENERAL INFORMATION:
; APPLICANT: Sinha, Sukanto
; APPLICANT: Seubert, Peter A.
; APPLICANT: Dovey, Harry F.
; APPLICANT: McConlogue, Lisa C.
; APPLICANT: Little, Sheila P.
; APPLICANT: Johnstone, Edward M.
; TITLE OF INVENTION: Amyloidin Protease and Uses Thereof
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: One Market Plaza, Steuart Street Tower,
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/059,032
; FILING DATE: 19930507
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, William E.
; REGISTRATION NUMBER: 30,223
; REFERENCE/DOCKET NUMBER: 15270-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 326-2400
; TELEFAX: (415) 326-2422
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 164 base pairs
```

TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-059-032-16

Query Match 2.0%; Score 41.4; DB 1; Length 164;  
Best Local Similarity 56.1%; Pred. No. 0.012;  
Matches 78; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
Qy 775 gaacacagaagatggaaatgcttttaatacaaggtgcaaaagaggaacacacata 834  
Db 26 GGAAGTGGAGGAGGCGCTTCAACTGCCGGTCCAGCGGTCGAAGGAGGAGAACTGCGCT 85  
Qy 835 atttgcagcagctactcccactgcgaacacaaagtgccaaactactcgtttatagaca 894  
Db 86 ATCTCAAGGAGCTGGTGACGCTGCGGGCCAGAAAGTCCCGCTGCTGGGTTCCACACG 145  
Qy 895 catgctgacttcgctcttg 913  
Db 146 CACGCCGACTATGCTCTGG 164

RESULT 12  
PCT-US91-07290-16  
Sequence 16, Application PC/YUS9107290  
GENERAL INFORMATION:  
APPLICANT: Sinha, Sukanto  
APPLICANT: Seubert, Peter A.  
APPLICANT: Dovey, Harry F.  
APPLICANT: McConlogue, Lisa C.  
APPLICANT: Little, Sheila P.  
APPLICANT: Johnstone, Edward M.  
TITLE OF INVENTION: Amyloidin Protease and Uses Thereof  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Athena Neurosciences, Inc.  
STREET: 800F Gateway Blvd.  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/07290  
FILING DATE: 19911004  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy, Lisabeth Feix  
REGISTRATION NUMBER: 31547  
REFERENCE/DOCKET NUMBER: 17796-002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 877-0900  
TELEFAX: (415) 877-8370  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 164 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
PCT-US91-07290-16

Query Match 2.0%; Score 41.4; DB 5; Length 164;  
Best Local Similarity 56.1%; Pred. No. 0.012;  
Matches 78; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
Qy 775 gaacacagaagatggaaatgcttttaatacaaggtgcaaaagaggaacacacata 834  
Db 26 GGAAGTGGAGGAGGCGCTTCAACTGCCGGTCCAGCGGTCGAAGGAGGAGAACTGCGCT 85  
Qy 835 atttgcagcagctactcccactgcgaacacaaagtgccaaactactcgtttatagaca 894  
Db 86 ATCTCAAGGAGCTGGTGACGCTGCGGGCCAGAAAGTCCCGCTGCTGGGTTCCACACG 145  
Qy 895 catgctgacttcgctcttg 913  
Db 146 CACGCCGACTATGCTCTGG 164  
RESULT 13  
US-09-056-105-23/c  
Sequence 23, Application US/09056105  
Patent No. 6287569  
GENERAL INFORMATION:  
APPLICANT: KIPPS, THOMAS J.  
APPLICANT: WU, YUNQI  
TITLE OF INVENTION: VACCINES WITH ENHANCED INTRACELLULAR  
TITLE OF INVENTION: PROCESSING  
FILE REFERENCE: 233/221  
CURRENT APPLICATION NUMBER: US/09/056.105  
CURRENT FILING DATE: 1998-04-06  
EARLIER APPLICATION NUMBER: 60/043,467  
EARLIER FILING DATE: 1997-04-10  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 23  
LENGTH: 1821  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-056-105-23

Query Match 1.7%; Score 36.2; DB 4; Length 1821;  
Best Local Similarity 55.0%; Pred. No. 1.3;  
Matches 71; Conservative 0; Mismatches 58; Indels 0; Gaps 0;  
Qy 486 cagacgatacttgaaagtcatttaaaatgggaaagaaatgggtccatcttcctga 545  
Db 249 CAGACGTTCTGAAAAAGACACACAGATCTCTGGGAAGGAACATCCCTATTTCTGG 190  
Qy 546 acaagtacagaatgaatcaatcaatgaagaaagaatgagtgatgtatgtattt 605  
Db 189 CATGGCACTGATTAGAAACAGTCAAAAAAAGAAAAAAGAAATATCCCTATT 130  
Qy 606 taacaaaaa 614  
Db 129 TAACCACAA 121  
RESULT 14  
US-08-253-503-1/c  
Sequence 1, Application US/08253503  
Patent No. 5589334  
GENERAL INFORMATION:  
APPLICANT: Coullie, Pierre  
APPLICANT: Boon-Falleur, Thierry  
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE WHICH CODES  
TITLE OF INVENTION: FOR A TUMOR REJECTION ANTIGEN PRECURSOR  
TITLE OF INVENTION: WHICH IS PROCESSED TO AN ANTIGEN PRESENTED  
TITLE OF INVENTION: BY HLA-B\*44, AND USES THEREOF  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Felfe & Lynch  
STREET: 805 Third Avenue  
CITY: New York

Query Match

	Matches	71: Conservative	0: Mismatches	58: Indels	0: Gaps
Qy	486	cagacgactctggaaaagtcaattaaatgggaaagaagaatgggtccatcttcctga	545		
Db	1198	CAGACGTTCTGAAAAGACACACAGATCTCTGGGAGGAACATCCCTATTTTCTGG	1139		
Qy	546	acaagtacagaatgaaatcgaatcgaatgaagaagaagaatgagtgcgtatgtattgattt	605		
Db	1138	CATGGCACTGATTGAAGAACAGTCAAAAAAATAAAAAAAGGAAATATCCCTATTT	1079		
Qy	606	taacaaaaa	614		
Db	1078	TAACCAAAA	1070		

```

RESULT 15
US-08-796-883-1/c
: Sequence 1, Application US/08796883
: Patent No. 5744353
: GENERAL INFORMATION:
: APPLICANT: Herman, Jean; Coulie, Pierre;
: APPLICANT: Boon-Fallier, Thierry; van der Bruggen, Pierre;
: APPLICANT: Luescher, Immanuel.
: TITLE OF INVENTION: Tumor Rejection Antigens Presented By
: TITLE OF INVENTION: HLA-B*44 Molecules, And Uses Thereof
: NUMBER OF SEQUENCES: 30
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Felfe & Lynch
: STREET: 805 Third Avenue
: CITY: New York City
: STATE: New York
: ZIP: 10022
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
: COMPUTER: IBM
: OPERATING SYSTEM: PC-DOS
: SOFTWARE: Wordperfect
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/796,883
: FILING DATE: 06-FEB-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/602,506

```

us-09-833-782-1.rni

Tue Jan 15 09:01:14 2002

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 18:56:36 ; Search time 63.23 Seconds  
(without alignments)  
408.225 Million cell updates/sec

Title: US-09-833-782-2

Perfect score: 3668

Sequence: 1 MIARCLLAVRRVGGSR.....FLKREPQKAFILMSRGLHAP 704

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3483	95.0	704	1 NEUL_PIG	Q02038 sus scrofa
2	3475	94.7	704	1 NEUL_RABIT	P42675 oryctolagus
3	3348	91.3	704	1 NEUL_RAT	P42676 rattus norv
4	2330	63.5	686	1 MEPE_PIG	P47788 sus scrofa
5	2330	63.5	686	1 MEPE_HUMAN	P52888 homo sapien
6	2291	62.5	686	1 MEPE_RAT	P24155 rattus norv
7	1085.5	29.6	712	1 PRVD_YEAST	P25375 saccharomyc
8	818.5	22.3	680	1 OPDA_ECOLI	P27298 escherichia
9	814.5	22.2	681	1 OPDA_HAEIN	P44573 haemophilus
10	806	22.0	680	1 OPDA_SALTY	P27237 salmoneilla
11	525	14.3	713	1 PMIP_HUMAN	Q99797 homo sapien
12	519	14.1	680	1 DCP_ECOLI	P24171 escherichia
13	502.5	13.7	680	1 DCP_SALTY	P27236 salmoneilla
14	472	12.9	710	1 PMIP_RAT	Q01992 rattus norv
15	461	12.6	772	1 PMIP_YEAST	P35999 saccharomyc
16	448.5	12.2	762	1 PMIP_SCHPO	Q10415 schizosacch
17	405.5	11.1	775	1 PMIP_SCHCO	P37932 schizophyll
18	145.5	4.0	3911	1 AK9_HUMAN	Q99996 h a kinase
19	133.5	3.6	590	1 PEFF_BORBU	O51264 borrelia bu
20	133	3.6	1341	1 RPAL_METVA	Q58445 methanococ
21	132	3.6	993	1 SCPL_MOUSE	Q62209 mus musculu
22	132	3.6	1427	1 REST_HUMAN	P30622 homo sapien
23	131.5	3.6	1818	1 T294_HUMAN	O94822 homo sapien
24	129.5	3.5	1539	1 Y373_HUMAN	O15078 homo sapien
25	129.5	3.5	3066	1 POLG_SBMVG	Q90069 s genome po
26	129	3.5	1972	1 MYHB_HUMAN	P35749 homo sapien
27	128	3.5	834	1 YN4_CAEEL	P34537 caenorhabdi
28	125	3.4	591	1 YN48_YEAST	P42846 saccharomyc
29	125	3.4	1235	1 DPOL_PYRHO	O58610 pyrococcus
30	124.5	3.4	5035	1 RYNR_PIG	P16960 sus scrofa
31	123.5	3.4	1087	1 AK9_RABIT	Q28628 oryctolagus
32	123	3.4	997	1 SCPL_RAT	Q03410 rattus norv
33	122.5	3.3	602	1 PEPE_LACLC	P94876 lactococcus

#### ALIGNMENTS

RESULT	NEUL_PIG	STANDARD;	PRT;	704 AA.
AC	Q02038:			
DT	01-OCT-1993 (Rel. 27, Created)			
DT	01-OCT-1993 (Rel. 27, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	NEULOLYSIN PRECURSOR (EC 3.4.24.16) (NEUTROTENSIN ENDOPEPTIDASE)			
DE	(MITOCHONDRIAL OLIGOPEPTIDASE M) (MICROSOMAL ENDOPEPTIDASE) (MEP)			
DE	(SOLUBLE ANGIOTENSIN-BINDING PROTEIN) (SABP) (ENDOPEPTIDASE 24.16).			
OS	Sus scrofa (Pig).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Eumalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.			
OX	NCBI_TaxID=9823;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RC	TISSUE=Heart, and Liver;			
RX	MEDLINE=92388174; PubMed=1517239;			
RA	Sugiura N., Hagiwara H., Hirose S.;			
RT	"Molecular cloning of porcine soluble angiotensin-binding protein.";			
RL	J. Biol. Chem. 267:18067-18072(1992).			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RX	MEDLINE=97326108; PubMed=9182559;			
RA	Kato A., Sugiura N., Saruta Y., Hosoiri T., Yasue H., Hirose S.;			
RT	"Targeting of endopeptidase 24.16 to different subcellular			
RT	compartments by alternative promoter usage.";			
RL	J. Biol. Chem. 272:15313-15322(1997).			
CC	-!- FUNCTION: BINDS ANGIOTENSINS AND ITS ANALOGS.			
CC	-!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE IN NEUTROTENSIN:			
CC	10-PRO- -TYR-11.			
CC	-!- COFACTOR: BINDS ONE ZINC ION.			
CC	-!- SUBCELLULAR LOCATION: MITOCHONDRIAL AND CYTOPLASMIC			
CC	(BY SIMILARITY).			
CC	-!- TISSUE SPECIFICITY: IS PREDOMINANTLY EXPRESSED IN THE LIVER,			
CC	KIDNEY AND ADRENAL GLAND.			
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);			
CC	ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; D11336; BAA01949.1;			
DR	EMBL; AB000170; BAA19060.1;			
DR	EMBL; AB000425; BAA19104.1;			
DR	EMBL; AB000411; BAA19104.1; JOINED.			
DR	EMBL; AB000414; BAA19104.1; JOINED.			
DR	EMBL; AB000415; BAA19104.1; JOINED.			
DR	EMBL; AB000416; BAA19104.1; JOINED.			
DR	EMBL; AB000417; BAA19104.1; JOINED.			

P35579 homo sapien  
P49454 homo sapien  
Q01397 neurospora  
Q9usi6 schizosacch  
P11047 homo sapien  
Q13439 homo sapien  
P02566 caenorhabdi  
P35748 oryctolagus  
P32380 saccharomyc  
P21817 homo sapien  
P79134 bos taurus  
P06529 bacillus su

```
DR EMBL; AB000418; BAAL19104.1; JOINED.
DR EMBL; AB000419; BAAL19104.1; JOINED.
DR EMBL; AB000420; BAAL19104.1; JOINED.
DR EMBL; AB000421; BAAL19104.1; JOINED.
DR EMBL; AB000422; BAAL19104.1; JOINED.
DR EMBL; AB000423; BAAL19104.1; JOINED.
DR EMBL; AB000424; BAAL19104.1; JOINED.
DR PIR; A43411; A43411.
DR MEROPS; M03.002; -.
DR InterPro; IPR001567; Peptidase_M3.
DR InterPro; IPR000130; Zn_MTPeptidse.
DR Pfam; PF01432; Peptidase_M3; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR TRANSIT; 1 37 MITOCHONDRION (BY SIMILARITY).
KW Metalloprotease; Hydrolase; Zinc; Mitochondrion; Transit peptide.
FT CHAIN 1 38 704 NEUROLYSIN.
FT METAL 497 497 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 498 498 BY SIMILARITY.
FT METAL 501 501 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 504 504 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 704 AA; 80756 MW; A02BPEC67B7044A1 CRC64;

Query Match 95.08; Score 3483; DB 1; Length 704;
Best Local Similarity 94.5%; Pred. No. 8.4e-208;
Matches 665; Conservative 15; Mismatches 24; Indels 0; Gaps 0;

Qy 1 MIARCLAVSLRVRGSGRIILRLMTLGRVMSPLQAMSSYTVAGRNVLRLDLSPEQIKR 60
Db 1 MIVRCLSAARLHRRVGGGILLRLMTLGRVMSPLQAMSSYTVAGRNVLRLDLSPEQIKR 60

Qy 61 TEELIVQTKQYDVGMLGIEVYENCLOALADVEVKYIVERTMLDPQHVSSDKEVRA 120
Db 61 TEELIAQTKQYDVGMLGIEVYENCLOALADVEVKYIVERTMLDPQHVSSDKEVRA 120

Qy 121 ASTEADKRLSRFDIEMSMRGDIFERIVHLQETCDLGKIKPEARRYLEKSIKMGKRNGLHL 180
Db 121 ASTEADKRLSRFDIEMSMRGDIFERIVHLQETCDLGKIKPEARRYLEKSIKMGKRNGLHL 180

Qy 181 PEQVQNETKSMKRMSELICIDFNKLNEDDFLVFSKAEIGALPDFDTSLEKTDGDKYK 240
Db 181 PEQVQNETKSMKRMSELICIDFNKLNEDDFLVFSKAEIGALPDFDTSLEKTDGDKYK 240

Qy 241 ITLKYPHYFPMVKCCIPETRRMEMAFNTRCKEENTIILOQLLPLRTKVAKLGYSTHA 300
Db 241 ITLKYPHYFPMVKCCIPETRRMEMAFNTRCKEENTIILOQLLPLRAKVAKLGYSTHA 300

Qy 301 DFVLEMTAKSTSRVTAFLDLSOKLPLGSAEREFILNLKKKECEKGFYDGGKINAWD 360
Db 301 DFVLEMTAKSTHVTAPFLDLSOKLPLGSAEREFILNLKKKECEKGFYDGGKINAWD 360

Qy 361 LYYVMTQTEELKYSIDQEFLEKYEPIEVVTEGLNTYQELGLSPEQMTDAHVNKSVTL 420
Db 361 LHYVMTQTEELKYSVDQETLKEYPEIEVVTTEGLNTYQELGLSFEQYTDHAVNKSVTL 420

Qy 421 YTVKDKATGVLGQFYLDLYPREGYNHAAFCGLQPGCLLPDGSRMMAVAALVNFSPQV 480
Db 421 YTVKDKATGVLGQFYLDLYPREGYNHAAFCGLQPGCLLPDGSRMMSVAALVNFSPQR 480

Qy 481 AGRSLLRHDEVRYTFHEFGVHMVQICAQTFARFSGTNVETDFVEVPSQMLENNWVDVD 540
Db 481 AGRPSLLRHDEVRYTFHEFGVHMVQICAQTFARFSGTNVETDFVEVPSQMLENNWVDTD 540

Qy 541 SLRLSKHYKDGSPITDLDLEKLVASRLVNTGLTLRQIVLSKYDOSLHTNTSLDAASEY 600
Db 541 SLRLSKHYKDGSPITDLDLEKLVASRLVNTGLTLRQIVLSKYDOSLHTNTSLDAASEY 600

Qy 601 AKYCEILGVAATPGTNPATFGHLAGGYDQYGYGLWSEFSDMFYSCFKKEGIMNPE 660
Db 601 AKYCEILGVAATPGTNPATFGHLAGGYDQYGYGLWSEFSDMFYSCFKKEGIMNPE 660

Qy 661 VGMKYRNLIKPGSLDGMQLONFLKREPQKAFMSRGLHAP 704
Db 661 VGMKYRNLIKPGSLDGMQLONFLKREPQKAFMSRGLHAP 704
```

```
Db 661 VGMKYRNLIKPGSLDGMQLONFLKREPQKAFMSRGLHAP 704

RESULT 2
NEUL_RABIT STANDARD; PRT; 704 AA.
AC P42675; 1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE NEUROLYSIN PRECURSOR (EC 3.4.24.16) (NEUROTENSIN ENDOPEPTIDASE)
DE (MITOCHONDRIAL OLIGOPEPTIDASE M) (MICROSOMAL ENDOPEPTIDASE) (MEP).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
[1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP SEQUENCE-LLiver;
RC MEDLINE=93286083; PubMed=8509389;
RX Kawabata S.-I., Nakagawa K., Muta T., Iwanaga S., Davie E.W.;
RA "Rabbit liver microsomal endopeptidase with substrate specificity for
RT processing proproteins is structurally related to rat testes
RT metalloendopeptidase 24.15."
RL J. Biol. Chem. 268:12498-12503(1993).
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE IN NEUROTENSIN:
CC 10-PRO-1-TYR-11.
CC -1- COFACTOR: BINDS ONE ZINC ION.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL AND CYTOPLASMIC
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);
CC ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D13310; BAA02570.1; -.
CC MEROPS; M03.002; -.
CC InterPro; IPR001567; Peptidase_M3.
CC InterPro; IPR000130; Zn_MTPeptidse.
CC Pfam; PF01432; Peptidase_M3; 1.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Metalloprotease; Hydrolase; Zinc; Mitochondrion; Transit peptide.
FT TRANSIT 1 37 MITOCHONDRION (BY SIMILARITY).
FT CHAIN 38 704 NEUROLYSIN.
FT METAL 497 497 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 498 498 BY SIMILARITY.
FT METAL 501 501 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 504 504 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 704 AA; 80689 MW; A1B7E4D3E38E8086C CRC64;

Query Match 94.7%; Score 3475; DB 1; Length 704;
Best Local Similarity 93.3%; Pred. No. 2.6e-207;
Matches 657; Conservative 30; Mismatches 17; Indels 0; Gaps 0;

Qy 1 MIARCLAVSLRVRGSGRIILRLMTLGRVMSPLQAMSSYTVAGRNVLRLDLSPEQIKR 60
Db 1 MIARCFSAVRGLHRRVGGSGRIILRLMTLGRVMSPLQAMSSYTVAGRNVLRLDLSPEQIKR 60

Qy 61 TEELIVQTKQYDVGMLGIEVYENCLOALADVEVKYIVERTMLDPQHVSSDKEVRA 120
Db 61 TEELIAQTKQYDVGMLGIEVYENCLOALADVEVKYIVERTMLDPQHVSTDEVRA 120

Qy 121 ASTEADKRLSRFDIEMSMRGDIFERIVHLQETCDLGKIKPEARRYLEKSIKMGKRNGLHL 180
Db 121 ASTEADKRLSRFDIEMSMRGDIFERIVHLQETCDLGKIKPEARRYLEKSIKMGKRNGLHL 180

Qy 121 ASTEADKRLSRFDIEMSMRGDIFERIVHLQETCDLGKIKPEARRYLEKSIKMGKRNGLHL 180
Db 121 ASTEADKRLSRFDIEMSMRGDIFERIVHLQETCDLGKIKPEARRYLEKSIKMGKRNGLHL 180
```



```

181 PEQVQNEIKSMKRMSELCDIDFNKNLNEDDTFLVFSKAELGALPDDFIDSLEKTDKDKYK 240
181 PEEVQNEIKSMKRMSELCDIDFNKNLNEDDTFLVFSKAELGALPDDFIDSLEKTDKDKYK 240
241 ITLKYPHYFPVVKKCCIPETRRMEMAFNTRCKEENTIILOQLLPLRTKAKLLGYSTHA 300
241 ITLKYPHYFPVVKKCCIPETRRMEMAFNTRCKEENTIILOQLLPLRTKAKLLGYSTHA 300
301 DFVLEMTAKTSRVTAFIDDLDSOKLKPGEAREFIRILNKKKECKDRGFEDYDGNKINAWD 360
301 DFVLEMTAKTSRVTAFIDDLDSOKLKPGEAREFIRILNKKKECKDRGFEDYDGNKINAWD 360
361 LYYMTQTEELKYSIDQEFKEFPIEVVTEGLNTYQELLGLSFGQMTDAHVWNKSVTL 420
361 LHYMTQTEELKYSIDQEFKEFPIEVVTEGLNTYQELLGLSFGQMTDAHVWNKSVTL 420
421 YTVKDKATGEVLGQFYLDLYPREGKYNHAACFGLQPCGLLPDGSRRMMAVAAVYVNSQPV 480
421 YTVKDKATGEVLGQFYLDLYPREGKYNHAACFGLQPCGLLPDGSRRMMAVAAVYVNSQPV 480
481 AGRPSLLRHDEVRTYFHEFGVHMHIQCAOTDFARFSGTNVETDFVEVPQMLNWWYDND 540
481 AGRPSLLRHDEVRTYFHEFGVHMHIQCAOTDFARFSGTNVETDFVEVPQMLNWWYDND 540
541 SURRLSKHYKDGSPADDDLEKLIVASRLVNTGLTLRQIVLSKVQDQSLHTNTSLDAASEY 600
541 SURRLSKHYKDGSPADDDLEKLIVASRLVNTGLTLRQIVLSKVQDQSLHTNTSLDAASEY 600
601 AKYCSILGVAATPGTNMPATFGLAGGYDQGYGYLWSEVFSMDMFYSCFKKEGIMNPE 660
601 ARYCTDILGVAATPGTNMPATFGLAGGYDQGYGYLWSEVFSMDMFYSCFKKEGIMNPE 660
661 VGMKYRNLIIRPGSLDGMMDLHNFILKREPNOKAFIUMSRGLHAP 704
661 VGMKYRNLIIRPGSLDGMMDLHNFILKREPNOKAFIUMSRGLHAP 704

RESULT 3
NEUL_RAT
ID NEUL_RAT STANDARD; PRT; 704 AA.
AC P42676;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE NEULOXIN PRECURSOR (EC 3.4.24.16) (NEUROTENSIN ENDOPEPTIDASE)
DE (MITOCHONDRIAL OLIGOPEPTIDASE M) (MICROSOMAL ENDOPEPTIDASE) (MEP).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RX MEDLINE=96070836; PubMed=7592986;
RA Dauch P., Vincent J.-P., Checler F.;
RT "Molecular cloning and expression of rat brain endopeptidase
RT 3.4.24.16.";
RL J. Biol. Chem. 270:27266-27271(1995).
RN [2]
RP SEQUENCE OF 38-57, AND CHARACTERIZATION.
RC TISSUE=Liver;
RX MEDLINE=95138171; PubMed=7836437;
RA Serizawa A., Dando P.M., Barrett A.J.;
RT "Characterization of a mitochondrial metalloproteinase reveals
RT neurolysin as a homologue of thimet oligopeptidase.";
RL J. Biol. Chem. 270:2092-2098(1995).
CC -!- FUNCTION: HYDROLYSES OLIGOPEPTIDES SUCH AS NEUROTENSIN,
CC BRADYKININ, DYNORPHIN A, ETC.
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE IN NEUROTENSIN:
CC 10-PRO-|-TYR-|-I.
CC -!- COFACTOR: BINDS ONE ZINC ION.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL INTERMEMBRANE SPACE AND
CC ALSO CYTOPLASMIC.

```

-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);  
ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.  
-----  
THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; X87157; CAA60630.1; -  
DR MEROPS; M03.002; -  
DR InterPro; IPR001567; Peptidase\_M3.  
DR InterPro; IPR000130; Zn\_MTPeptidse.  
DR Pfam; PF01432; Peptidase\_M3; 1.  
DR PROSITE; PS001442; ZINC\_PROTEASE; 1.  
KW Metalloprotease; Hydrolase; Zinc; Mitochondrion; Transit peptide.  
FT TRANSIT 1 37  
FT CHAIN 38 704  
FT ACT\_SITE 498 498  
FT METAL 497 497  
FT METAL 501 501  
FT METAL 504 504  
SQ SEQUENCE 704 AA; 80253 MW; E33F7967A79343D1 CRC64;  
  
Query Match 91.3%; Score 3348; DB 1; Length 704;  
Best Local Similarity 89.7%; Pred. No. 1.8e-199;  
Matches 630; Conservative 40; Mismatches 32; Indels 0; Gaps 0;  
  
QY 1 MIARCLAVRSRRVGGSRILLRMTLGRVMSPLQAMSSYTVAGRNVLRLWDLSPQLIKR 60  
DB 1 MITLCLSTLRGLHRAGGSRLLQLTMTLGELASPLQAMSSYTAAGRNVLRLWDLSPQLIKR 60  
QY 61 TEELIVQTKQVYDAVGMGLGIEEVTYENCLQALADVEVKYIVERTMLDFFQHVSSDREVA 120  
DB 61 TEQLIAQTKQVYDVTGCTIALKEVTYENCLQVLADEVTYIVERTMLDFFQHVSSDREVA 120  
QY 121 ASTEADKRLSRFDIEMSRMGDIFERIVHLOETCDLQKIKPEARRYLEKSIKMGKRNGLHL 180  
DB 121 ASTEADKLSRFDIEMSRMGDVFQIRVHLOETCDLEKIKPEARRYLEKSIKMGKRNGLHL 180  
QY 181 PEQVQNEIKSMKRMSELCDIDFNKNLNEDDTFLVFSKAELGALPDDFIDSLEKTDKDKYK 240  
DB 181 SEHIREIKSMKRMSELCDIDFNKNLNEDDTSLVFSKAELGALPDDFIDSLEKTDKDKYK 240  
QY 241 ITLKYPHYFPVVKKCCIPETRRMEMAFNTRCKEENTIILOQLLPLRTKAKLLGYSTHA 300  
DB 241 VTLKYPHYFPVVKKCCVPETRRKMEMAFNTRCKQENTAILQQLPLRAQVAKLLGYNTHA 300  
QY 301 DFVLEMTAKTSRVTAFIDDLDSOKLKPGEAREFIRILNKKKECKDRGFEDYDGNKINAWD 360  
DB 301 DFVLEMTAKTSRVAADFDDLSQKLPGEAREFIRILNKKKECKDRGFEDYDGNKINAWD 360  
QY 361 LYYMTQTEELKYSIDQEFKEFPIEVVTEGLNTYQELLGLSFGQMTDAHVWNKSVTL 420  
DB 361 LHYMTQTEELKYSVDQESLKEFPIEVVTEGLLSIYQELLGLSFGQVDAHVWNKSVSL 420  
QY 421 YTVKDKATGEVLGQFYLDLYPREGKYNHAACFGLQPCGLLPDGSRRMMAVAAVYVNSQPV 480  
DB 421 YTVKDKATGEVLGQFYLDLYPREGKYNHAACFGLQPCGLLPDGSRRMMAVAAVYVNSQPV 480  
QY 481 AGRPSLLRHDEVRTYFHEFGVHMHIQCAOTDFARFSGTNVETDFVEVPQMLNWWYDND 540  
DB 481 AGRPSLLRHDEVRTYFHEFGVHMHIQCAOTDFARFSGTNVETDFVEVPQMLNWWYDND 540  
QY 541 SURRLSKHYKDGSPADDDLEKLIVASRLVNTGLTLRQIVLSKVQDQSLHTNTSLDAASEY 600  
DB 541 SURRLSKHYKDGSPADDDLEKLIVASRLVNTGLTLRQIVLSKVQDQSLHTNTSLDAASEY 600  
QY 601 AKYCSILGVAATPGTNMPATFGLAGGYDQGYGYLWSEVFSMDMFYSCFKKEGIMNPE 660  
DB 601 AKYCSILGVAATPGTNMPATFGLAGGYDQGYGYLWSEVFSMDMFYSCFKKEGIMNPE 660

Db 601 AKYTEILGVAATPGTNMPATFGHLAGGYDQYGYUWSEVFSMDPHSCFKKKGIMNPE 660  
QY 661 VGMKYNLILKPGSLDGMFLNFKREPQKRAFLMSRGLH 702  
DB 661 VGMKYNLILKPGSLDGMFLNFKREPQKRAFLMSRGLN 702  
RESULT 4  
MEPD\_PIG STANDARD; PRT; 686 AA.  
AC P47786;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE THIMET OLIGOPEPTIDASE (EC 3.4.24.15) (ENDOPEPTIDASE 24.15).  
GN THOPI.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=94222057; PubMed=8168506;  
RA Kato A., Sugliura N., Hagiwara H., Hirose S.;  
RT "Cloning, amino acid sequence and tissue distribution of porcine  
RT thimet oligopeptidase. A comparison with soluble angiotensin-binding  
RT protein.";  
RL Eur. J. Biochem. 221:159-165 (1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=97326108; PubMed=9182559;  
RA Kato A., Sugliura N., Saruta Y., Hosoiri T., Yasue H., Hirose S.;  
RT "Targeting of endopeptidase 24.16 to different subcellular  
RT compartments by alternative promoter usage.";  
RT J. Biol. Chem. 272:15313-15322 (1997).  
CC -1- FUNCTION: INVOLVED IN THE METABOLISM OF NEUROPEPTIDES UNDER 20  
CC AMINO ACID RESIDUES LONG. INVOLVED IN CYTOPLASMIC PEPTIDE  
CC DEGRADATION  
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF THE PHE-|-ALA BOND  
CC IN SUBSTRATES CONTAINING -PHE-PHE-|-ALA-ALA-PHE- AND SIMILAR  
CC SEQUENCES.  
CC -1- COFACTOR: BINDS ONE ZINC ION.  
CC -1- SUBUNIT: MONOMER.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -1- TISSUE SPECIFICITY: UBIQUITOUS.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);  
CC ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D21871; BAA04882.1; -  
DR EMBL; AB000438; BAA19107.1; -  
DR EMBL; AB000426; BAA19107.1; JOINED.  
DR EMBL; AB000427; BAA19107.1; JOINED.  
DR EMBL; AB000428; BAA19107.1; JOINED.  
DR EMBL; AB000429; BAA19107.1; JOINED.  
DR EMBL; AB000430; BAA19107.1; JOINED.  
DR EMBL; AB000431; BAA19107.1; JOINED.  
DR EMBL; AB000432; BAA19107.1; JOINED.  
DR EMBL; AB000433; BAA19107.1; JOINED.  
DR EMBL; AB000434; BAA19107.1; JOINED.  
DR EMBL; AB000435; BAA19107.1; JOINED.  
DR EMBL; AB000436; BAA19107.1; JOINED.  
DR EMBL; AB000437; BAA19107.1; JOINED.  
DR MEROPS; M03.001; -

DR InterPro: IPR001567; Peptidase\_M3.  
DR InterPro: IPR000130; Zn\_Mrpeptidse.  
DR Pfam: PF01432; Peptidase\_M3; 1.  
DR PROSITE: PS00142; ZINC\_PROTEASE; 1.  
KW Metalloprotease; Hydrolase; zinc.  
FT INIT\_MET 0 BY SIMILARITY.  
FT METAL 472 472 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 473 473 BY SIMILARITY.  
FT METAL 476 476 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 479 479 ZINC (CATALYTIC) (BY SIMILARITY).  
SQ SEQUENCE 686 AA; 78013 MW; AA6006223047580C CRC64;  
Query Match 63.5%; Score 2330; DB 1; Length 686;  
Best Local Similarity 64.0%; Pred. No. 1.3e-136;  
Matches 431; Conservative 104; Mismatches 130; Indels 8; Gaps 1;  
QY 29 EVMSPLQAMSSYTVAGRNVLWDLSPOIKRTREELIVOTKQVYDAVCMGLGIEEVTENC 88  
DB 12 DVAAPCSAY-----NYLRWDLSSAQIGELTTTELEQTKRVYDRVGTQELQDVSYENT 63  
QY 89 LQALADVEVKYIVERTMLDPHQVSSDKVEVRAASTEADKRLSRFDIEMSMRGDIFERIVH 148  
DB 64 LKALADVEVSYTVQNRNLDLPQHVSPCKDITASTADKLSSEDFVMSMQDYVQRIWV 123  
QY 149 LQETCDLGGKIKPEARRYLEKSIMKGRNGLHLPQVONEIKSMKRMSELICIDFNKNLNE 208  
DB 124 LQEKVQKDSLRLPEARYLERLIKLRNGLHLPKETQEKIKSIKKLSLLCIDFNKNLNE 183  
QY 209 DDTFLVFSKAEALGALPDFFDSLEKTDKVKITLKYPHYPPVMKKCCIPETRRMEMAF 268  
DB 184 DTTFLPVTREELGGLPEDFLNSLEKTEDEKLUKVTLPKPHYFPLLKCHVPETRRKVEAF 243  
QY 269 NTRCKEENTILQOLLPLRTKVAKLGYSTHADFLVLENTAKTSRVTAFDLDLSQKLKP 328  
DB 244 NCRCKENCALIRELRLRAQKSSILGTFSTHADVLENNAKTSQVVTATFDELAQKLKP 303  
QY 329 LGEAREFILNKKKECKDRGFYDGGKINAWDLIYYTQTTELKYSIDQEFLEKYEPIEV 388  
DB 304 LGEQERAVILELKAECTKRGDLDFGRINAWDMRYMNQVETRYRVDONLLKYEFPQV 363  
QY 389 VTEGLLNTYQELLGLSPEQMTDAHWNKSVTLTYTKDKATGEVLGQFYLDLYPREGKYNH 448  
DB 364 VTRGLLIYQELLGLTFHLEGAAVHEDVMLYSVDRDAASKVIGKGYLDLYPREGKYGH 423  
QY 449 AACFGLQPGCLLPDGRSMMAVAALVNF SQPVAGRPSLLRHDEVRTYFHEFGHVMHQICA 508  
DB 424 AACFGLQPGCLRQDGRSQIAIAAMVANFTKPTDPAPSLQLQHDEVETYFHEFGHVMHQLCS 483  
QY 509 QTDFAFSGTNVETDVEVPQSMLENVWVDVSLRLSKHYKDGSPADLLEKLVASRL 568  
DB 484 QAEEFAMFSGTHVERDFVEAPSQMLENWWYEAEPILRMSQHYRTGSAIPEOLELKLKRSQ 543  
QY 569 VNTGLLTQLQVLKQDQSLHTNTSLDAASEYKVCSEILGVAATPGTNMPATFGHLAGG 628  
DB 544 ANTGLFNLRQVLAKVDQALHTQTAADPAEYARLCQELGVPATPGTNMPATFGHLAGG 603  
QY 629 YDGYGYGLWSEVFSMDMFYSCFKKEGIMNPEVGMKYNLILKPGSLDGMFLNFKR 688  
DB 604 YDAQYGYLWSEVYSADMFHTREFQEGILSGKVGMDYRSCILRPGGSEDAVYMLKFLGR 663  
QY 689 EPNOKAFMLMSRGL 701  
DB 664 DPKODAFLLSKGL 676  
RESULT 5  
MEPD\_HUMAN STANDARD; PRT; 688 AA.  
ID MEPD\_HUMAN  
AC P52888;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE THIMET OLIGOPEPTIDASE (EC 3.4.24.15) (ENDOPEPTIDASE 24.15) (MP78).  
 GN THOP1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=95367027; PubMed=7639763;  
 RA Thompson A., Huber G., Walther P.;  
 RT "Cloning and functional expression of a metalloendopeptidase from  
 human brain with the ability to cleave a beta-APP substrate  
 peptide.";  
 RL Biochem. Biophys. Res. Commun. 213:66-73(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Taylor G.R., Otulakowski G., Lau C.Y., Munroe D.G.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: INVOLVED IN THE METABOLISM OF NEUROPEPTIDES UNDER 20  
 CC AMINO ACID RESIDUES LONG. INVOLVED IN CYTOPLASMIC PEPTIDE  
 CC DEGRADATION. ABLE TO DEGRADE THE BETA-AMYLOID PRECURSOR PROTEIN  
 CC AND GENERATE AMYLOIDGENIC FRAGMENTS.  
 CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF THE PHE-|-ALA BOND  
 CC IN SUBSTRATES CONTAINING -PHE-PHE-|-ALA-ALA-PHE- AND SIMILAR  
 CC SEQUENCES.  
 CC -!- COFACTOR: BINDS ONE ZINC ION.  
 CC -!- SUBUNIT: MONOMER (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);  
 CC ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 -----  
 DR EMBL; 250115; CA90477.1; -;  
 DR EMBL; 029366; AAA82607.1; -;  
 DR MEMOPS; M03.001; -;  
 DR MIM; 601117; -;  
 DR InterPro; IPR001567; Peptidase M3.  
 DR InterPro; IPR001130; Zn\_MTPeptidse.  
 DR Pfam; PF01432; Peptidase\_M3; 1.  
 DR PROSITE; P500142; ZINC\_PROTEASE; 1.  
 KW Metalloprotease; Hydrolase; Zinc.  
 FT INIT\_MET 0 BY SIMILARITY.  
 FT METAL 472 472 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 473 473 BY SIMILARITY.  
 FT METAL 476 476 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 479 479 ZINC (CATALYTIC) (BY SIMILARITY).  
 SQ SEQUENCE 688 AA; 78708 MW; 25ACF9139CD7EDF CRC64;

Query Match 63.5%; Score 2330; DB 1; Length 688;  
 Best Local Similarity 64.9%; Pred. No. 1.3e-136;  
 Matches 426; Conservative 97; Mismatches 133; Indels 0; Gaps 0;

QY 46 NVLRWDLSPQIKTRTEELIVOTKQYDVGMLGTEEVYENCLOALADVVKYIVERTM 105  
 DB 21 NDLRWLSAQAEERTELEIQTQKRYDVGQTEFEDSYESTLKALADVETVTVQRM 80  
 QY 106 LDFPQHVSDDKVRAASTEADKRLRFDIEMSMRGDIFERIVHLOETCDLGKIKPEARY 165  
 DB 81 LDFPQHVSFKDIRASTEADKRLSEEDVEMSMREDVYQRIVWLQEKVKQDSLRPEARY 140  
 QY 166 LKSKTMGRNGLHLPEQVQNEIKSKRMSELCIDFNKNLNEDDTFLVFSKAEGLGALPD 225  
 DB 141 LERLIKLRNGLHLPRETQENIKRIKKKLSLLCIDFNKNLNEDDTFLPFTLQELGGLPE 200

QY 226 DFIDSLEKTDKKYKITLKYPHYFFPMKCCIPETRRRMMEMAFNTRCKEENTIIQQQLP 285  
 DB 201 DFLNSLEKMEDCKLKVTLKYPHYFFPLKKCHVPETRRKVEEAFNCRCKEENCAILKELVT 260  
 QY 286 LRTKVALLLGYSTHADFLVLEMTAKTSRVTAFDLDLSOKLPLGEAREFRLNLKKKEC 345  
 DB 261 LRAQKSRLLGFTHADYVLEMMNAKTSQVATFLDELAKQLKPLGQERAVILELKRAC 320  
 QY 346 KDRGFYDGGKINAWDLYYMYTQTEELKYSIDOEFLEKPEPIEVVTEGLLNTYOELLGLSF 405  
 DB 321 ERGLPFDGRIIRANDMYMNOVEETRYCDQNLKEYFPVQVYVTHGLLGIQELLGLAF 380  
 QY 406 EQMTDAHWNKSVTLTYVKDKATGEVLGQFYLDLYPREGKYNHAACFGLQPCCLPDGSR 465  
 DB 381 HHEGASAWHEDVRLYTARDAASGEVVGKLYLDLYPREGKYGHAACFGLQPCCLRDGSR 440  
 QY 466 MNAVAAALVYVNSOPVAGRPSSLLRHDEVRYFHEFGVHMHOICAOITDFARFSGTNETDFV 525  
 DB 441 QIAIAAMVANFTKPTADAPSLQHDDEVETFFHEFGVHMHQLCSQAEFAMFSGTHVERDFV 500  
 QY 526 EYPSOMLENWVDVSLRLSKHYKDGSPIDDLLEKLVASRLVNTGLTLRQIVLSKYD 585  
 DB 501 EAPSOMLENWVWEQPLLRMSRHYRTGSAVPRELLEKLTESQANTGLNLRQIVLAKYD 560  
 QY 586 QSLHTNTSLDAASEYAKYCEILGVAATPGTNMPATFGHLAGGYDQGYGYLWSEYFSMD 645  
 DB 561 QALHTQTDADPAEYARLQCEILGVPATPGTNMPATFGHLAGGYDAQYGYLWSEYFSMD 620  
 QY 646 MYSYCFKKEGINPEVGMKYRNLIKPGSLDGMMLHNFLEKREPQKAFLSRGL 701  
 DB 621 MFHTRFKQEGVLNKGMDYRCILLRPGGSEDASAMLRFLRFLGRDPKQDAFLLSKGL 676

RESULT 6  
 MEPD\_RAT STANDARD; PRT; 686 AA.  
 ID P24155;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE THIMET OLIGOPEPTIDASE (EC 3.4.24.15) (ENDO-OLIGOPEPTIDASE A)  
 DE (ENDOPEPTIDASE 24.15) (PZ-PEPTIDASE) (SOLUBLE METALLO-ENDOPEPTIDASE).  
 GN THOP1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Testis;  
 RX MEDLINE=91084500; PubMed=2261476;  
 RA Pierotti A., Dong K.-W., Glucksman M.J., Orłowski M., Roberts J.L.;  
 RT "Molecular cloning and primary structure of rat testes  
 metalloendopeptidase EC 3.4.24.15.";  
 RL Biochemistry 29:10323-10329(1990).  
 RN [2]  
 RP REVISIONS.  
 RX MEDLINE=94029935; PubMed=8216239;  
 RA McKie N., Dando P.M., Rawlings N.D., Barrett A.J.;  
 RT "Thimet oligopeptidase: similarity to 'soluble angiotensin II-binding  
 protein' and some corrections to the published amino acid sequence of  
 the rat testis enzyme.";  
 RL Biochem. J. 295:57-60(1993).  
 CC -!- FUNCTION: INVOLVED IN THE METABOLISM OF NEUROPEPTIDES UNDER 20  
 CC AMINO ACID RESIDUES LONG. INVOLVED IN CYTOPLASMIC PEPTIDE  
 CC DEGRADATION.  
 CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF THE PHE-|-ALA BOND  
 CC IN SUBSTRATES CONTAINING -PHE-PHE-|-ALA-ALA-PHE- AND SIMILAR  
 CC SEQUENCES.  
 CC -!- COFACTOR: BINDS ONE ZINC ION.  
 CC -!- SUBUNIT: MONOMER.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -!- TISSUE SPECIFICITY: IS EXPRESSED ABUNDANTLY IN THE TESTIS.

IT IS ALSO FOUND IN THE LIVER, LUNG AND KIDNEY.  
-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);  
ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M61142; AAA41586.1; -  
PIR; A36165; HYRTH.  
PIR; S38760; S38760.  
MEROPS; M03.001; -  
InterPro; IPR001567; Peptidase\_M3.  
InterPro; IPR00130; Zn\_Mtpeptidse.  
Pfam; PF01432; Peptidase\_M3; 1.  
PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW Metalloprotease; Hydrolase; Zinc.  
INIT\_MET 0  
METAL\_ 472 473 ZINC (CATALYTIC) (BY SIMILARITY).  
ACT\_SITE 473 476 BY SIMILARITY.  
METAL\_ 476 476 ZINC (CATALYTIC) (BY SIMILARITY).  
METAL\_ 479 479 ZINC (CATALYTIC) (BY SIMILARITY).  
SEQUENCE 686 AA; 78183 MW; F50E629D0E0B6EA12 CRC64;

Query Match 62.5%; Score 2291; DB 1; Length 686;  
Best Local Similarity 64.3%; Pred. No. 3,3e-134;  
Matches 422; Conservative 98; Mismatches 136; Indels 0; Gaps 0;

QY 46 NVLRWLSPEOIKTEBTELVQTVQVDAVGMGLTEEVYENCLOADADVVKYIVERTM 105  
DB 21 NHLRWLSAQOIRALTQIIEQTCVDRVGAQDFEDVSYLETKALADVEVTVQVQNI 80  
QY 106 LDFPHVSDSKEVRAASTADKRLSRFDIEMSRGDIPIERVHJQETCDLKGIRPEARRY 165  
DB 81 LDFPHVSPKNDIRAASTEADKRLSEFDVEMSRQDVQYRVVWLQETPKDLSKPEARRY 140  
QY 166 LKSTKMKCRNGLHPEOVQNEIKSMKRMSELCLDFNKNLNEEDTFLVFSKALGALPD 225  
DB 141 LERLKLGRNGLHPDQTPQKIRNIKKLSLLCLDFNKNLNEEDTFLVFSKALGALPE 200  
QY 226 DFIDSELTDDKVKITLKYPHYPPVMKCCIPETRRMEMAFNTRCKEENTIILOQLLP 285  
DB 201 DFLNLEKTEDEGKLVTKLYPHYPLPKKCHVPETRRLLLEAFNCRCKEENCAILKELVS 260  
QY 286 LRTKVALGLYSTHADFLVEMNTAKTSRVTAFDLDLSQKLKPLGEAREFILLNKKEC 345  
DB 261 LRAQSNLGLFRTHADYVLENNMAKTSQVATFDELARKLPLGEQERAVILELKEAS 320  
QY 346 KDRFEYDCKINAWDLYYIMQTEBELKYSIDQEFKVEFPVIEVTEGLLNTYQELLGLSF 405  
DB 321 AKRGLPFDGRHAMDYRMVNOVEDSVYDQNLKKEYFPQVTVRGLLAIYQELLGLTF 380  
QY 406 EQMTDAHVNKSVTLTYVKDKATGEVLQGVLDLYPRGKYNHAAFCGLQPCGLLPDQSR 465  
DB 381 TLEEGAAWHEDVRLYSVRDAASGEIEGKFDYLDLYPRGKYGHACFGCLQPCGLRQDSR 440  
QY 466 MMAVALVYNSQPVAGRSLLRHDEVTYFHEFGHVMHQICAQPDFARFGTNNVETDFV 525  
DB 441 QLAIAAMVANFTKPTPDVPSLLQHDVEVTFHEFGHVMHQLCSQAEPAMFSGTHVERDFV 500  
QY 526 EVPQSMLENWVDVSLRRLSKHKYGDSPIADDDLLLEKLVASRLVNTGLTLRQIVLSKVD 585  
DB 501 EAPQSMLENWVEKEPLMRMSQHYRTGGEAPDELEKLIKRSQAGNAGLNLRIIVLAKVD 560  
QY 586 QSLHTNTSLDAASEYAKYCSSETLGAATPGTNMPATFGHLAGGYDQYGYLWSEVFSMD 645  
DB 561 QVLHTQTDVDAEYARLCCQELGLVGPATPGTNMPATFGHLAGGYDAQYGYLWSEVFSMD 620

QY 646 MFYSCFKKEGIMPEVGMKYRNLLIKPGGSLDGMMLHNFLLKREPNOKAFILMSRGL 701  
DB 621 MFHTRFQEGVLSPKVGMDYRTSILRPGGSEDASTMLKQLGRLDPKQDAFLLSKGL 676

## RESULT 7

ID PRD\_YEAST STANDARD; PRT; 712 AA.  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 13-DEC-1998 (Rel. 37, Last annotation update)  
DE SACCCHAROLYSIN (EC 3.4.24.37) (PROTEASE D) (PROTEINASE YSCD)  
GN (OLIGOPEPTIDASE YSCD).  
OS PRD1 OR YCL057W OR YCL57W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94139744; PubMed=8307027;  
RA Buechler M., Tislar U., Wolf D.H.;  
RT "Proteinase yscd (oligopeptidase yscd). Structure, function and  
RT relationship of the yeast enzyme with mammalian thimet oligopeptidase  
RT (metalloendopeptidase, EP 24.15).";  
RL Eur. J. Biochem. 219:627-639(1994).  
RN [2]  
RP SEQUENCE OF 1-663 FROM N.A.  
RA Alexandraki D., Galland P., Thireos G., Tzermia M.;  
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 663-712 FROM N.A.  
RA Voet M., Volckaert G.;  
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: COULD BE INVOLVED IN LATE STAGE OF PROTEIN DEGRADATION.  
CC -1- CATALYTIC ACTIVITY: CLEAVAGE OF PRO-|-PHE AND ALA-|-ALA BONDS.  
CC -1- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);  
CC ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X76504; CAA54039.1; -  
EMBL; X59720; CAA42388.1; -  
PIR; S19387; S19387.  
MEROPS; M03.003; -  
SGD; S000562; PRD1.  
InterPro; IPR001567; Peptidase\_M3.  
InterPro; IPR00130; Zn\_Mtpeptidse.  
Pfam; PF01432; Peptidase\_M3; 1.  
PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW Hydrolase; Metalloprotease; Zinc.  
FT METAL 501 501 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 502 505 BY SIMILARITY.  
FT METAL 505 505 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 508 508 ZINC (CATALYTIC) (BY SIMILARITY).  
SEQUENCE 712 AA; 81933 MW; 340910B7FDAFBE37 CRC64;

Query Match 29.6%; Score 1085.5; DB 1; Length 712;  
Best Local Similarity 35.2%; Pred. No. 9.5e-60;  
Matches 251; Conservative 130; Mismatches 295; Indels 37; Gaps 13;

QY 20 ILLRWTLGKREYNSPL-----QAMSSYTVAGRNVLRLWDLSPQIKTRTEELIVQTKQ 70  
DB 20 ILLRWTLGKREYNSPL-----QAMSSYTVAGRNVLRLWDLSPQIKTRTEELIVQTKQ 70





```

AC DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE OLIGOPEPTIDASE A (EC 3.4.24.70).
GN PRLC OR OPDA OR OPTA.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-18.
RX MEDLINE=92165739; PubMed=1537805;
RA Conlin C.A., Miller C.G.;
RT "Cloning and nucleotide sequence of opda, the gene encoding
RT oligopeptidase A in Salmonella typhimurium.";
RL J. Bacteriol. 174:1631-1640(1992).
CC -!- FUNCTION: MAY PLAY A SPECIFIC ROLE IN THE DEGRADATION OF SIGNAL
CC PEPTIDES AFTER THEY ARE RELEASED FROM PRECURSOR FORMS OF SECRETED
CC PROTEINS. CAN CLEAVE N-ACETYL-L-ALA4.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF OLIGOPEPTIDES, WITH BROAD
CC SPECIFICITY. GLY OR ALA COMMONLY OCCUR AS P1 OR P1' RESIDUES, BUT
CC MORE DISTANT RESIDUES ARE ALSO IMPORTANT. AS IS SHOWN BY THE FACT
CC THAT 2-GLY-PRO-GLY-1-GLY-PRO-ALA IS CLEAVED, BUT NOT 2-(GLY)5.
CC -!- COFACTOR: BINDS ONE ZINC ION.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);
CC ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M84574; AAA27172.1; -
DR PIR; A42298; A42298.
DR MEROPS; M03.004; -.
DR StyGene; SG10311; opda.
DR InterPro; IPR001567; Peptidase_M3.
DR InterPro; IPR000130; Zn_Mtpeptdse.
DR Pfam; PF01432; Peptidase_M3; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR Hydrolase; Metalloprotease; Zinc.
FT METAL 469 469 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 470 470 BY SIMILARITY.
FT METAL 473 473 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 476 476 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 680 AA; 76943 NW; CA0F11C80PE13F89 CRC64;
Query Match 22.0%; Score 806; DB 1; Length 680;
Best Local Similarity 30.4%; Pred. No. 1.6e-42;
Matches 206; Conservative 129; Mismatches 300; Indels 42; Gaps 17;
QY 52 LSPEQIKTRTEELIVTKQYDAVGMGLTEETVENCLOALADVEKVIYVERTMLDFPOH 111
Db 17 IRPEHVVPATKALACRAAEVGVAHG-APYSWENLCQPLA--EADDVLGRFSI-SH 72
QY 112 VSSDK---EVRAASTADKRLSRFDIEMSRMGDIFERIVHLQETCDLGKIKPPEARYLEK 168
Db 73 LNSVKNSPELREAYEQTLPLLSSEYSTWVGHEGLYNAYRDLRDSGDHYATINTAQKKAVDN 132
QY 169 SIKMGKRNGLHLPQEQVQNEIKSMKRMSELCTDFKNLNEDD---TFLVFSKAEALGALPD 225
Db 133 ALRDFELSGIGLPKQRYGEIATRLSELNGQYSNNVLDATMGWTKLITDEAELAGMPE 192
QY 226 DFIDS---LEKTDKDYKITLKYPPYFPMKKCCIPETRRRMEAFNTRC-----K 273
Db 193 SALAAKAQAEKEQGYLLTLDIPSLPVMYCDNQALREMYRAYTSRQDQGNAGK 252
QY 274 EENTIILQOLLPLRTKVAKLGLGYSTHADEVLEMNTAKSTSRVTAFLDLDLSQKLKPLGEAE 333

```

```

Db 253 WDNSPVMEEILALRHELAQLLGFENVAHESLATKMAENPQOVLDFLTDLAKRARPOGEKE 312
QY 334 REFILNLKKKECKDRGFYDVGKINAWDLYYMTQTTELKYSIDQEFKKEYFPIEVTEGL 393
Db 313 LAQLRAFAKAEF---GVE---ELQPWDIAIYSEKQHLIYSISDEQLRYPFENKAVNGL 366
QY 394 LNTYQELLGLSFEQMTDAHWKNSVTLTYTKDKATCEVLGQFVLDLYPREGKYNHA---A 450
Db 367 FEVVKRIYGITAKERTDQVWHPEVRFELYDE--NNELRGSFYLDLYAREHKGGAMDD 425
QY 451 CFGLQPGCLLPDGSRMMAVAALVNFSPQVACRPSLLRHDEVRYEFHFCHVMHQCAQT 510
Db 426 CVGQMRKA--DGTLOKPVAYLTNCFNRPVNGKPAFTHTDEVITLFEHFGHLLHMLTRI 482
QY 511 DPARFSG-TNVEDTFVEVFSQMLENVWVDLSRLSKYHKDGSPIADDDLEKLVASRLV 569
Db 483 ETAGVSGISGVWDVAVELPSQFMENWCPEALAFSTGHYETGEPLPKELLDKMLAAKNY 542
QY 570 NTGLLTRLQVLVLSKVDQSLTHTNTSLDAASEYAKYCSFI-LGVAATPCT---NMPATFGHL 625
Db 543 QAALFLRLQLEGLDFRLHAENPQOGAKILETLEIKKQVAVVSPTWGRPHAFSHI 602
QY 626 -AGGYDQGYGYLWSEVFSMDMFYSCFKKEGIMNPVGMKRYRLILKPGSGSLDGMDLHN 684
Db 603 FAGGYAAGYVYVLWADVLAADA-YSRFEEBEGFNRETGQSLDNLITRGSGSEEPMEFKR 661
QY 685 FLKREPQKAFMLSRGL 701
Db 662 FRGREPQLDAMLEHYGI 678
RESULT 11
PMIP_HUMAN
ID PMIP_HUMAN STANDARD; PRT; 713 AA.
AC Q99797;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE MITOCHONDRIAL INTERMEDIATE PEPTIDASE PRECURSOR (EC 3.4.24.59) (MIP).
GN MIP-OR MIP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97230465; PubMed=9073519;
RA Chew A., Buck E.A., Peretz S., Sirugo G., Rinaldo P., Isaya G.;
RT "Cloning, expression, and chromosomal assignment of the human
RT mitochondrial intermediate peptidase gene (MIP-OR).";
RL Genomics 40:493-496(1997).
CC -!- FUNCTION: CLEAVES PROTEINS, IMPORTED INTO THE MITOCHONDRION, TO
CC THEIR MATURE SIZE.
CC -!- CATALYTIC ACTIVITY: RELEASE OF AN N-TERMINAL OCTAPEPTIDE AS SECOND
CC STAGE OF PROCESSING OF SOME PROTEINS IMPORTED IN THE
CC MITOCHONDRION.
CC -!- ENZYME REGULATION: ACTIVITY IS DIVALENT CATION-DEPENDENT. IT IS
CC STIMULATED BY MANGANESE, MAGNESIUM OR CALCIUM IONS AND REVERSIBLY
CC INHIBITED BY ZINC, COBALT, AND IRON (BY SIMILARITY).
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);
CC ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```







```

DR PIR; A49931; A49931.
DR MEROPS; M03.005; -.
DR EcoGene; EG10212; dcp.
DR InterPro; IPR001567; Peptidase_M3.
DR InterPro; IPR000130; Zn_Mtpeptidase.
DR Pfam; PF01432; Peptidase_M3; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Dipeptidase;
KW Calcium; Complete proteome.
FT INIT_MET 0
FT METAL 469
FT ACT_SITE 470
FT METAL 473
FT METAL 476
FT CONFLICT 138
FT SEQUENCE 680 AA; 77384 MW; 04B9DCC03E15893A CRC64;

Query Match 14.1%; Score 519; DB 1; Length 680;
Best Local Similarity 25.7%; Pred. No. 8.6e-25;
Matches 157; Conservative 120; Mismatches 299; Indels 34; Gaps 12;

QY 114 SKEVRASTEADKRLSRDIEMSRGDIPIRVHQLQETCDLGKIKPEARVYLEKSMKG 173
DB 83 TNDLQRLDEQFSAEALAEANDIYLGELFARVDVWQRRESGLDSEIRLVEIHQRF 142
QY 174 KRGLHLPEOVONEIKSMKRMSELICIDFNKML--NEDDTLIVFSKALGALPDDFI-- 228
DB 143 VLAKAKLAQADKAKLVLTAEATTSQFNQRLAANKSGGLVNDIOLAGMSEQEIATL 202
QY 229 ---DSLEKDDDKYKITLKYPHYFVPMKKCCIPETRRMEMAFNTRCKE---ENTIIQL 281
DB 203 AAFAAREKGLDNKWLPLNTTQOPALAEWRDRATEKLFIAGTWRAEKNDANDTRAIQ 262
QY 282 QLLPLRTKVKILGYSTHADFLVLENNATKSTSRVTAFLDLSQKLPGLGEAREFILTNIK 341
DB 263 RLVEIRAQOATLILGPHYAAKIAQADOMAKTPEAALNFMREIVPAARQASDELASIQAVI 322
QY 342 KKECKDRGFEYDGNKINAWDLYXYMTQTEELKYSIDOEFLKEYFPTE-VYTEGLLNTYQEL 400
DB 323 DKQ--OGGF---SAQPDWAFYAEQVRREKFDLDLDAQKPYFELNTVLNEGVTANQL 376
QY 401 LGLSFEQMTDAHWKNSVTLTVKDKATGEVLGQFVLDLYPREGKYNHAAFCGLOPGCLL 460
DB 377 FGKIFVEREDIPYHPDVRVWEIFDH-NGVGLALFGDFEARDKSG-----GAWMGNFV 430
QY 461 PDC--SRMVAALVNFVQPVAGRPSLLRHDEVRTYFHEFGVHMHIQCAQTDFAFSGT 518
DB 431 EQSTLNKTHPIVYVNCYQKPAAGEPALLLDVITLPHFEGHTLHGLFARQRYATLSGT 490
QY 519 NVETDFVEVPSQMLENVWDVDSRLSKHYKDGSPDIADDLLEKLVASRLVNTGLTLRQ 578
DB 491 NTRDVEFPQSQINHEWATHPQVFARYAHYOSGAMPELOQKMRNASLFNKGEMSEL 550
QY 579 IVLSKVDQSLHNTSLDAAS-----EYAKYCSFELGVAATPGTNMPATGHL-AGGYDQ 632
DB 551 LSAALLDMWHCLEENAEQDQVDDFELRALVAENMDLPAIPRYSSYFAHIFGGGYAG 610
QY 633 YGYLWSEVFSMDMFYCPKKGIMNPEVGMKYRNILKPGSGLDGMDMLHNLKREPNO 692
DB 611 YYAYLWTOQLA--DGGYQWFEVQGLTRENGRLFRFRAILSRGNSDELERLYRQWRGKAPKI 669
QY 693 KAFLSRGLH 702
DB 670 MPMLQHRGLN 679

RESULT 13
DCP_SALTY
ID DCP_SALTY STANDARD; PRT; 680 AA.
AC P27236;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)

```

```

DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PEPTIDYL-DIPEPTIDASE DCP (EC 3.4.15.5) (DIPEPTIDYL CARBOXYPEPTIDASE).
GN DCP.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-12.
RX MEDLINE=92165738; PubMed=1537804;
RA Miller C.G., Hamilton S.;
RT "Cloning and nucleotide sequence of the Salmonella typhimurium dcp
RT gene encoding dipeptidyl carboxypeptidase.";
RL J. Bacteriol. 174:1626-1630(1992).
CC -!- FUNCTION: REMOVES DIPEPTIDES FROM THE C-TERMINI OF N-BLOCKED
CC TRIPEPTIDES, TETRAPEPTIDES AND LARGER PEPTIDES.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF UNBLOCKED, C-TERMINAL DIPEPTIDES
CC FROM OLIGOPEPTIDES, WITH BROAD SPECIFICITY. DOES NOT HYDROLYSE
CC BONDS IN WHICH PL' IS PRO, OR BOTH PL AND PL' ARE GLY.
CC -!- COFACTOR: BINDS ONE ZINC ION.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);
CC ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M84575; AAA27055.1; -.
DR PIR; A42297; A42297.
DR MEROPS; M03.005; -.
DR StyGene; SG10083; dcp.
DR InterPro; IPR001567; Peptidase_M3.
DR InterPro; IPR000130; Zn_Mtpeptidase.
DR Pfam; PF01432; Peptidase_M3; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Dipeptidase.
FT METAL 469
FT ACT_SITE 470
FT METAL 473
FT METAL 476
FT SEQUENCE 680 AA; 77261 MW; DDFD89B47B660B45 CRC64;

Query Match 13.7%; Score 502.5; DB 1; Length 680;
Best Local Similarity 26.8%; Pred. No. 9e-24;
Matches 190; Conservative 115; Mismatches 302; Indels 103; Gaps 24;

QY 61 TPEELIVQTKQVYDAVGMGLGIEVTVENCILQ---ALADVEVKYIVE-----RTMLDPP 109
DB 3 TNPLLDQSLPYPAPRFDRIKDCYRPAFDEGVQRKRVIEAIVNHVPAAPDFNTLLALE 62
QY 110 QH---VSSDKVRASTEAA--DKRLSRFDIEMS-----MRGDIPIRVHQL-QET 152
DB 63 QSGALLSRVTSVFFAMTAHTNDELQRLDEAFSAELAAALSNDIYLSALFARVDVWQOR 122
QY 153 CDLGKIKPEARVYLEKSIKMGKRNGLHLPEOVONEIKSMKRMSELICIDFNKMLNEDDTF 212
DB 123 HSLG-LDDESRLVDVIHORFVLAGAQLAEEDKARLKVLTNTESATLMSFNQRL----- 175
QY 213 LVFSKALGALPDDFIDSL-----EKTDDDKYKITLKYPHYFVPMVKCC 256
DB 176 LAASKA--GGLAVDDAHCLAGLSPEEMTVAAEAAREKGLEERWFILPNTTQOPALATLR 233
QY 257 IPETRRRMEAFNTRCKE-----ENTIIQLPLRTKVKAKLLGYSTHADFLVLENNATKST 312
DB 234 DRQTRNLFAASWTRAERKGDADHTRAIVQRLVEIRRCQAKLLGFPNYAAWKADQAKTP 293
QY 313 SRVTAFLDLSQKLPGLGEAREFILTNIKKECKDRGFEYDQK-----INAWDLYXYMTQ 367

```

Db 294 QAAALSF-----MRGIVPPARQORVLN-EQAEIQN-----VIDGEQGYTVQAWDMFYAEQ 342  
QY 368 TEEKYSIDQEEFKEXEPIE-VVTEGLNLYQELGLSFEQMTDAHWNKSVTLTYVKDK 426  
Db 343 VRREKALDEAQLKPYFALNTVLOEGFWTANQIFGTFIVERDEIPVHPDVRVWEIFD- 401  
QY 427 ATGEVLQGFYLDLYPREGKYNHAAFCGLQGLCLPDG--SRMMAVAALVNFSPVAGR 484  
Db 402 SDGVGMALFYCDFARSKSG-----CAWGNFVEQSTLNETRPVIYNVCNYQKPVQDGP 456  
QY 485 SLLRHDEVRYTFHEFGHVMHQICAQTFARFSGTNVETDFVEVPSQMLENWNVDVSLRR 544  
Db 457 ALLLWDDVITLFHEFGHTLGLFAVQRYATLSGTNTRDFVEFVPSQINEHWASPRVFER 516  
QY 545 LSKHYKGSPIADLLEKLVASRLVNTG-----LLTLQIVL--SKVDQSLHTNT 592  
Db 517 YARVHDEGKMPADQERMKRKLNFNGYDWTLLGAALDMRWMLLEESVAEQSV----- 572  
QY 593 SLDAASVYAKYCSILGVAATPGTNMPATFGLH-AGGYDGYGYGLWSEVFSMDMFYSCF 651  
Db 573 --REFEQOALAAEHLDPVAPPRYSYFAHIFGGYAGYAYLWTQMLA-DDGYQWF 628  
QY 652 KKEGIMPEVGMKYNILILPGSLDGMMLHNLKREPNOKAFILMSRGL 701  
Db 629 VEQGLTRENGORFRDAILARGNSTDLTLYSAWRGHEPHIDPMLQYRGL 678  
RESULT 14  
PMIP\_RAT  
ID PMIP\_RAT STANDARD; PRT; 710 AA.  
AC Q01992;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE MITOCHONDRIAL INTERMEDIATE PEPTIDASE PRECURSOR (EC 3.4.24.59) (MIP).  
GN MIPEP OR MIP.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OC NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=92390438; PubMed=1518864;  
RA Isaya G., Kalousek F., Rosenberg L.E.;  
RT "Sequence analysis of rat mitochondrial intermediate peptidase:  
RT similarity to zinc metallopeptidases and to a putative yeast  
RT homologue.";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:8317-8321(1992).  
RN [2]  
RP SEQUENCE OF 34-50 AND 528-542.  
RC TISSUE=Liver;  
RX MEDLINE=92347312; PubMed=1322290;  
RA Kalousek F., Isaya G., Rosenberg L.E.;  
RT "Rat liver mitochondrial intermediate peptidase (MIP): purification  
RT and initial characterization.";  
RL EMBO J. 11:2803-2809(1992).  
CC -1- FUNCTION: CLEAVES PROTEINS, IMPORTED INTO THE MITOCHONDRION, TO  
CC THEIR MATURE SIZE.  
CC -1- CATALYTIC ACTIVITY: RELEASE OF AN N-TERMINAL OCTAPEPTIDE AS SECOND  
CC STAGE OF PROCESSING OF SOME PROTEINS IMPORTED IN THE  
CC MITOCHONDRION.  
CC -1- ENZYME REGULATION: ACTIVITY IS DIVALENT CATION-DEPENDENT. IT IS  
CC STIMULATED BY MANGANESE, MAGNESIUM OR CALCIUM IONS AND REVERSIBLY  
CC INHIBITED BY ZINC, COBALT, AND IRON.  
CC -1- SUBUNIT: MONOMER.  
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);  
CC ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation

the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M96633; AAA1899.1; --  
DR PIR; A46273; A46273.  
DR MEROPS; M03.006; --  
DR InterPro; IPR001567; Peptidase\_M3.  
DR InterPro; IPR000130; Zn\_MTpeptidse.  
DR Pfam; PF01432; Peptidase\_M3.1  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW Hydroxylase; Metalloprotease; Zinc; Transit peptide; Mitochondrion;  
KW Magnesium; Manganese; Calcium; Cobalt; Iron.  
FT CHAIN 1 33 MITOCHONDRION.  
FT METAL 34 710 MITOCHONDRIAL INTERMEDIATE PEPTIDASE.  
FT ACT\_SITE 492 492 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 493 493 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 496 496 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 499 499 ZINC (CATALYTIC) (BY SIMILARITY).  
SQ SEQUENCE 710 AA; 80673 MW; 4FC2E6743EA43558 CRC64;  
  
Query Match. 12.9%; Score 472; DB 1; Length 710;  
Best Local Similarity 24.4%; Pred. No. 7.3e-22;  
Matches 187; Conservative 119; Mismatches 294; Indels 166; Gaps 28;  
  
QY 17 GSRILLRMTLGRVMSPLQAMSSYTVAGRNVL- - - - -WDL--- 52  
Db 6 GTRYAYKL-CGRRAAALQGR- - - - -AGRCARSVSWSVPVGAFAFNKPGQHLWLLGE 59  
QY 53 -----SPEIKTRTEELIVOTKQYDAV- - - - -GMLGIEEVTYENCALQA 93  
Db 60 RRGFGVPPELSTPEGFQVQAEALRKTELVERACSTPPGPQTVLIEDELS--DCLCRVA 117  
QY 94 DVEVKYIVERTMLDFPOHVSSDKEVRAASTFADKRLSRFDIEMSMRGDIERIVHLOETC 153  
Db 118 D-----LADFVKIGHPEAFREAQAEACISGTVMVEKLTNTNVELYQ--SLOKLL 164  
QY 154 D---LGKIKPEARRYLEKSIKMGKRNGLHLPQEQVQNEIKSMKKRMSLCLIDFN-KNLNE 208  
Db 165 DDKKLMSLDAETRRVAELPFWDFEISGIHLDEB- - - - -KRRRA--VDLVNKILDL 213  
QY 209 DDTFLVFS---KAELGALPDDF- - - - -ISL--EKTDD---DKYKITLKP 246  
Db 214 SSAFLMGNTNPKIKQKHLPEHIQHFFARDGRHLVDGLHAEASDDLVRFAAYKIFL-Y 272  
QY 247 HYFPVMKCCIPETRRRMEMAFNTRCKEENTIILOQLPLRTKVKAKLLGYSTHADFVLEM 306  
Db 273 N-----ADQLKC-----LEELLSSDLLANLVGY-----LPF 299  
QY 307 NT-----AKSTSRVTAFLDLSQKLPLGEAREFILLNKKKECKDRGFYDGKINAWD 360  
Db 300 PTGPPGTAOTPETVMQFLEKLSKLCERTKDFEMQGMKTKLNPO-----NSELAPWD 354  
QY 361 LYYMTOTEELKYSIDQEEFLKEYPIEVVTEGLNTYQELLGLSF--EQMTDAHWKNSV 418  
Db 355 PPYSYGVIRAERYNIEPSLYCPFLSLGACMEGLNVLENRLGLVTLAEQPFKEGVWCIDV 414  
QY 419 TLYTVKDKATGEVLGQFVLDLYPREGKYNHAAFCGLQGLCLPDGSRMMAVAALVNFVSQ 478  
Db 415 RKLAVVHESEG-LLGYIYCDFQFRANKPQQDCHTTRGRUKEDGSGYQLPVVLMNLPH 473  
QY 479 PVAGRPSILLRHDEVRYTFHEFGHVMHQICAQTFARFSGTNVETDFVEVPSQMLENWNVD 538  
Db 474 ASRDFPILLPGMMENLFHEGHAMSHMLGTRYQHVHTGTCTPTDFAEVPISILMEYSND 533  
QY 539 VDSLRRLSKHYKGSPIADLLEKLVASRLVNTGTLRLQIVLQVLSKVDQSLHTNTSLDAS 598  
Db 534 YRVVSQFAKHVYQTGPLPKAMVSRCLCESKVKCAAAEQLOVFAALDQIYHGQPL----- 589  
QY 599 EYAKYCSIL-----GVAATPTNTNPTATFGLAGGYDGYGYGLWSEVFSMDMFYSC 650

Db 590 --KKSTDIILMETEQYGLPYVPTAWQLRFSHLV-GYGAKYISYLSRAVSMWVWKEC 646

QY 651 FKKEGIMNEVGMKYNILILKPGGSLDGLDMLHNFILKREPKNQAF 696

Db 647 FLODP-FNRAAGERYRRELAHGGKREPLMTIQMLQKPCSIDDF 691

RESULT 15

PMIP\_YEAST STANDARD; PRT; 772 AA.

AC P35999; P51980;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE MITOCHONDRIAL INTERMEDIATE PEPTIDASE PRECURSOR (EC 3.4.24.59) (MIP).

GN OCT1 OR MIP1 OR YKL134C.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI\_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=94309680; PubMed=8035833;

RA Isaya G., Miklos D., Rollins R.A.;

RT "MIP1, a new yeast gene homologous to the rat mitochondrial intermediate peptidase gene, is required for oxidative metabolism in Saccharomyces cerevisiae.";

RT Mol. Cell. Biol. 14:5603-5616(1994).

RL [2]

RN SEQUENCE FROM N.A.

RA Rad M.R., Xu G., Kirchrath L., Fritz C., Keuchel H., Hollenberg C.P.;

RL Submitted (MAR-1994) to the EMBL/Genbank/DBJ databases

CC -!- FUNCTION: CLEAVES PROTEINS, IMPORTED INTO THE MITOCHONDRION, TO THEIR NATURAL SIZE. CLEAVES THE NUCLEAR-ENCODED PRECURSORS FOR CYTOCHROME OXIDASE SUBUNIT IV (COXIV) AND THE IRON-SULFUR PROTEIN (FE-S) OF THE BC1 COMPLEX.

CC -!- CATALYTIC ACTIVITY: RELEASE OF AN N-TERMINAL OCTAPEPTIDE AS SECOND STAGE OF PROCESSING OF SOME PROTEINS IMPORTED IN THE MITOCHONDRION.

CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE); ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U10243; AAA21278.1; -

DR EMBL; Z28134; CAAB1975.1; -

DR PIR; S37963; S37963.

DR MEROPS: M03.006; -

DR SGD; S0001617; OCT1.

DR InterPro; IPR001567; Peptidase\_M3.

DR InterPro; IPR001030; zn\_Mtpeptdse.

DR Pfam; PF01432; Peptidase\_M3; 1.

DR PROSITE; PS00142; ZINC\_PROTEASE; 1.

KW Hydrolase; Metalloprotease; Zinc; Transit peptide; Mitochondrion.

FT TRANSIT 1 37

FT CHAIN 38 772 MITOCHONDRIAL (POTENTIAL).

FT METAL 558 558 MITOCHONDRIAL INTERMEDIATE PEPTIDASE.

FT ACT\_SITE 559 559 BY SIMILARITY.

FT METAL 562 562 ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 565 565 ZINC (CATALYTIC) (BY SIMILARITY).

FT CONFLICT 343 350 MAKNPADV -> WQDRRC (IN REF. 2).

FT CONFLICT 694 694 Y -> S (IN REF. 2).

FT CONFLICT 696 702 APYYSYL -> QLITATY (IN REF. 2).

FT CONFLICT 702 702

FT SEQUENCE 772 AA; 88182 MW; 1C19A0655FAAE7CA CRC64;

Query Match 12.6%; Score 461; DB 1; Length 772;

Best Local Similarity 25.6%; Pred No. 3.9e-21;

Matches 176; Conservative 127; Mismatches 291; Indels 94; Gaps 29;

QY 53 SPEQIKTRTEELIVQTKQYVDVAVGMLGIE--EVTYENCIALQALADVEKYIVERTMLDFPQ 110

Db 85 SPDGLRKFSQVSLQAQAEILLDKMRNDFSESGKLTITMNLDRSLTLCR-----VIDLCE 138

QY 111 HVSS-----DKEYRAASTEADKRLSRFDIEMSRGDIIFERIVHLOETCDLGIK---RPEAR 163

Db 139 FIRSTHPDDAFVRAAQ---DCHQEMFEFNVNLNTDV-----SLCNILKSVLNNPEVS 187

QY 164 RYLE-KSIRMGK-----RNGHLPEQVQNETKSMKRMSELCIDFNKLNLEDDEDFLVF 215

Db 188 SKLSAEELKVGKILLDDPEKSGIYMPDPVREKFIQISQEISLVGQEE---INHTD---Y 240

QY 216 SKAELGALPDFDIDS-----LEKTD---KYKI-TLKYPHYPPVMKKCCIPETRRR 263

Db 241 PGSNSVYKIPCKDLNDSKVSFTFLKQLNKDVKGQNYKVPFTFGYAA-ALLKSCENEMVRKK 299

QY 264 MEMAFNTRCKEENTIILOQLPLRTKVKALLGYSTHADFVLEMTAKTSRVTAFDLDLS 323

Db 300 LMTALHS-CSDKQVKRSLHLIKLRAILANLMHKTSYAEYQLEGKMAKNPKDVQDFILTM 358

QY 324 QKLKPLGEAREERFILNKKKECK---DRGFEYDGTINAWDLVYVYMTQTEEL--KYSIDQ 377

Db 359 NNTIETANELKFAELKAKDLKPLTNTNDEILKLVRPWRDRYYGKYFQLNPSNPN 418

QY 378 EFLKEYFFIEVTEGLNTYQELLGLSFE-QMTD-AHVWNKSVTLTYVKDKATGEVLGQF 435

Db 419 KEISYFTLGNVIQGLSDFLQOIYIGIRLEPAITDEGETWSPDVRRLNVISEEG-IIGII 477

QY 436 YLDLYPREGKYNHAACF-----GLQPGCLLPDGSR-MMVAVALVNFQ 478

Db 478 YCDLFRNGKTSNPAHFTVCCSRQIYPSETDFSTIQVG-ENPDGTYFQLPVLVLCNFS 536

QY 479 -PVAGRPSL--LRHDEVRTYFHEFGVHMOICQATDFARFSGTNVETDFVEVPQMLENW 535

Db 537 ILTASKSKLCFLQISEVETLFEHMGHAMSLGRTMONGISGTRCATDFVELPSILMEHF 596

QY 536 VMDVDSILRLSKHYKDGSPITADDLLEKLVASRLVNTGLTLRQIVLSKVDQSLHTN--- 591

Db 597 AKDIRILITKIGHYGTETIQADMLQREMKSTNFLQNCETYSQAKWMLDQSFHDEKII 656

QY 592 --TSLDAASEYAKYKCEILGVAATPGTNMPATPGHLAGGYDQYGYGLMSEVFSMDMFYS 649

Db 657 DIDNFDVVENY-QALERRLKVLVDQSNMCGRFGHLF-GYGATYYSYLFORTIASKIWA 714

QY 650 CFKKEGIMNPEVGMKYRNILILKPGGSILD 677

Db 715 LFEDDPYSRKN-GDKFKKHLKLLKRWGLKD 741

Search completed: January 14, 2002, 19:03:34

Job time: 418 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 18:44:31 ; Search time 64.93 Seconds  
(without alignments)  
825.918 Million cell updates/sec

Title: US-09-833-782-2  
Perfect score: 3668  
Sequence: 1 MIARCLLAVRSLRRVGSRI.....FLKREPQKAFMLRGLHAP 704  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3483	95.0	704	2 A43411	soluble angiotensin
2	3475	94.7	704	2 A43985	microsomal endopep
3	2330	63.5	689	1 HYHUTH	thimet oligopeptid
4	2330	63.5	689	1 HYHUTH	thimet oligopeptid
5	2279	62.1	687	1 HYRTHH	thimet oligopeptid
6	1085.5	29.6	712	2 S19387	saccharolysin (EC
7	848	23.1	680	2 C82353	oligopeptidase A V
8	818.5	22.3	680	2 S47718	oligopeptidase A (
9	817.5	22.3	680	2 B86021	oligopeptidase A (
10	814.5	22.2	681	2 C64055	thimet oligopeptid
11	806	22.0	680	2 A42298	thimet oligopeptid
12	794.5	21.7	681	2 E83636	oligopeptidase A P
13	702	19.1	678	2 B81224	oligopeptidase A N
14	694	18.9	678	2 G81996	oligopeptidase A (
15	690.5	18.8	713	2 S76766	hypothetical prote
16	671.5	18.3	685	2 F75370	oligopeptidase A -
17	635.5	17.3	723	2 T49985	oligopeptidase A X
18	627.5	17.1	674	2 F82844	oligopeptidase A X
19	578.5	15.8	716	2 E82620	peptidyl-di-peptida
20	520	14.2	681	2 D85720	dipeptidyl carboxy
21	519	14.1	681	2 E64908	peptidyl-di-peptida
22	502.5	13.7	680	2 A42297	peptidyl-di-peptida
23	495.5	13.5	678	2 T02867	peptidyl dipeptida
24	472	12.9	710	1 A46273	mitochondrial inte
25	448.5	12.2	762	2 T38081	probable mitochond
26	415	11.3	770	2 S37963	mitochondrial inte
27	268.5	7.3	265	2 F37271	metalloproteinase
28	267.5	7.3	268	2 F37271	metalloproteinase
29	267	7.3	290	2 E37271	metalloproteinase

hypothetical prote  
probable peptidase  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
conserved hypothet  
rhoptry protein -  
hypothetical prote  
conserved hypothet  
oligopeptidase - H  
polyprotein allerg  
oligopeptidase  
hypothetical prote  
probable coiled-co  
DNA-directed RNA p  
microtubule-vesicl

30 189 5.2 820 2 T27909  
31 175.5 4.8 615 2 F82132  
32 145.5 4.0 1922 2 T00637  
33 143 3.9 532 2 C83696  
34 140 3.8 2166 2 G70163  
35 139.5 3.8 852 2 D72230  
36 139.5 3.8 2269 2 T28677  
37 139 3.8 1084 2 G71329  
38 137.5 3.7 880 2 F75103  
39 135 3.7 578 2 B71934  
40 135 3.7 1095 2 T31423  
41 133.5 3.6 590 2 H70130  
42 133 3.6 547 2 H72311  
43 133 3.6 1115 2 T41342  
44 133 3.6 1345 2 A64430  
45 133 3.6 1392 2 A43336

ALIGNMENTS

RESULT 1  
A43411  
soluble angiotensin-binding protein - pig  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 20-Jun-2000  
C:Accession: A43411  
R:Sugiura, N.; Hagiwara, H.; Hirose, S.  
J. Biol. Chem. 267, 18067-18072, 1992  
A:Title: Molecular cloning of porcine soluble angiotensin-binding protein.  
A:Note: sequence extracted from NCBI backbone (NCBIN:112884, NCBIP:112886)  
A:Reference number: A43411; MUID:92388174  
A:Accession: A43411  
A:Molecule type: mRNA  
A:Residues: 1-704 <SUG>  
A:Cross-references: GB:D11336; NID:g217708; PIDN:BAA01949.1; PID:g217709  
A:Note: sequence extracted from NCBI backbone (NCBIN:112884, NCBIP:112886)  
C:Superfamily: thimet oligopeptidase  
C:Keywords: mitochondrion

Query Match 95.0%; Score 3483; DB 2; Length 704;  
Best Local Similarity 94.5%; Pred. NO. 1.5e-203;  
Matches 665; Conservative 15; Mismatches 24; Indels 0; Gaps 0;

QY 1 MIARCLLAVRSLRRVGSRIILLRMTLGREVMSPLOAMSSYTVAGRNVLRLWDLSPQIKTR 60  
DB 1 MIYRCLSAARLHRVGSGLILLRMTLGREVMSPLOAMSSYTVAGRNVLRLWDLSPQIKTR 60  
QY 61 TEELIVOTKQYDAVGMLGIEEVTYENCLQALADVEVKYIVERTMLDFFQHVSSDKEVRA 120  
DB 61 TEELIAQTKQYDDIGMLDIEEVTYENCLQALADVEVKYIVERTMLDFFQHVSSDKEVRA 120  
QY 121 ASPEADKRISRFDIEMSRGDIIFERIVHQTCDLGKIKPEARRYLEKSIKMKRNLGHL 180  
DB 121 ASPEADKRISRFDIEMSRGDIIFERIVHQTCDLGKIKPEARRYLEKSIKMKRNLGHL 180  
QY 181 PEQVQNEIKSMKRMSELCLDFNKNLNEEDTFLVFSKAELGALPDDFIDSLKTDQDKYK 240  
DB 181 PEQVQNEIKSMKRMSELCLDFNKNLNEEDTFLVFSKAELGALPDDFIDSLKTDQDKYK 240  
QY 241 ITLKYPHYFVPMKKCCIPETRRRWMEMAFNTRCKEENTIIQLQLPLRTKVKAKLIGYSTHA 300  
DB 241 ITLKYPHYFVPMKKCCIPETRRRWMEMAFNTRCKEENTIIQLQLPLRTKVKAKLIGYSTHA 300  
QY 301 DFVLEMTAKTSRVTAFDLDLSQKLKPLGEAREFETLNLKKKECKDRGFEGYDGNKINAWD 360  
DB 301 DFVLEMTAKTSRVTAFDLDLSQKLKPLGEAREFETLNLKKKECKDRGFEGYDGNKINAWD 360  
QY 361 LYYMTQTTELKYSIDQEFLEKYPFIEVTEGLLNYQELLGLISFEQMTDAHVNNKSVTL 420  
DB 361 LHYMTQTTELKYSVDQEFLEKYPFIEVTEGLLNYQELLGLISFEQMTDAHVNNKSVTL 420  
QY 421 YTVKDKATGEVLQGFYLDLYPREGKYNHAAACFGLQPCCLLPDGSRMMAVAALVYVNFSPV 480



Db 425 AACFGLOPCLRODGRQIAIAAWANFTKPTDPASLLQHDDEVETFFHEFGVHMQLCS 484

QY 509 QTFARFSGTNGTDEVEVPSQMLENNWVDVSLRLSKHYKDGSPDIADLLLEKLVASRL 568

Db 485 QAEEAFSGTHVERDVEAPSOMLENNWVEAPLLRMSQHYRTGSAIPQELLEKLKRSQ 544

QY 569 VNTGLTLRQIVLSKYDQSLHTNTSLDAASEYAKYCEILGVAATPGTNMPATFGHLAG 628

Db 545 ANTGLPNLRQIVLAKVDQALHTQTAAAPAEYARLCQELILGVPATPGTNMPATFGHLAG 604

QY 629 YDQYYGYLWSEVFSMDMYSCFKKGINPEVGMKYRNLLILKPGSLDGMMLHNLFR 688

Db 605 YDQYYGYLWSEVFSMDMYSCFKKGINPEVGMKYRNLLILKPGSLDGMMLHNLFR 664

QY 689 EPNOKAFMLSRGL 701

Db 665 DPQDAFLSKGL 677

RESULT 4

HYHUTH

N:Alternate names: beta-amyloid precursor protein processing metalloproteinase; metalloproteinase

C:Species: Homo sapiens (man)

C:Date: 27-Sep-1995 #sequence\_revision 17-Nov-1995 #text\_change 18-Jun-1999

C:Accession: J04197; PC4053; G01881; G01882; A53633

R:Thompson, A.; Huber, G.; Malherbe, P.

Biochem. Biophys. Res. Commun. 213, 66-73, 1995

A:Title: Cloning and functional expression of a metalloendopeptidase from human brain with

A:Reference number: J04197; MUID:95367027

A:Accession: J04197

A:Molecule type: mRNA

A:Residues: 1-689 <THO1>

A:CROSS-references: GB:Z50115; NID:g1030054; PIDN:CAA90477.1; PID:g1030055

A:Experimental source: brain

A:Molecule type: protein

A:Accession: PC4053

A:Residues: 67-78;181-200 <THO2>

R:Munroe, D.G.

submitted to the EMBL Data Library, June 1995

A:Reference number: H00694

A:Accession: G01881

A>Status: translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-689 <MUN1>

A:CROSS-references: EMBL:U29366; NID:g1098599; PIDN:AAA82607.1; PID:g1098600

R:Munroe, D.G.

submitted to the EMBL Data Library, June 1995

A:Reference number: H00695

A:Accession: G01882

A>Status: translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-5 <MUN2>

A:CROSS-references: EMBL:U29367; NID:g1098597; PIDN:AAA82606.1; PID:g1098598

R:Papadimitris, G.; Siman, R.; Scott, R.; Abraham, C.R.

Biochemistry 33, 192-199, 1994

A:Title: Identification of a metalloprotease from Alzheimer's disease brain able to degrade

A:Reference number: A53633; MUID:94114484

A:Accession: A53633

A:Molecule type: protein

A:Residues: 67-78;181-197,'X',199-200 <PAP>

A:Experimental source: brain of Alzheimer's disease patient

A>Note: sequence modified after extraction from NCBI backbone

C:Comment: This metalloproteinase is responsible for cleaving a Met-Asp bond in the Alzheimer's

C:Genetics:

A:Gene: GDB:THOP1; GDB:ME78

A:CROSS-references: GDB:595011; GDB:642232; OMIM:601117

A:Map position: 19q13.3-19q13.3

C:Function:

A:Description: endopeptidase preferentially hydrolyzing peptide bonds on the carboxyl side

C:Superfamily: thimet oligopeptidase

C:Keywords: Alzheimer's disease; amyloid; brain; glycoprotein; hydrolase; metalloproteinase

F:2-689/Product: thimet oligopeptidase #status predicted <MAT>

F:451/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:473,477,480/Binding site: zinc (His) #status predicted

F:474/Active site: Glu #status predicted

Query Match 63.5%; Score 2330; DB 1; Length 689;

Best Local Similarity 64.9%; Pred. No. 1.1e-133;

Matches 426; Conservative 97; Mismatches 133; Indels 0; Gaps 0;

QY 46 NVLRWDLSPQIKTRTEELIVQTKQYDVGMLGIEBVEYENCLOALADVEVYIVERTM 105

Db 22 NDLRWDLSPAQKIEERTRELIEQTQKRYDQTOFEFDSVESTLKALADVEVYIVQNR 81

QY 106 LDPQHVSSDKVEVRAASTADKRLSRFDLEMSMRGDIFFERIVHLQETCDLGKTKPEARY 165

Db 82 LDFPQHVSPSKDJRTASTADKRLSEDFVEMSMREDYQRIWVLEQKVKQDSLRPEARY 141

QY 166 LEKSTRKMGKRNGLHLPEQVQNEIKSMKRMSELCDIFNKNLNEDDTFLVFSKAEALGALPD 225

Db 142 LERLIKLRNRNGLHLPRETQENIKRIKKLLSLCIDFNKNLNEDTFLPFTLQELGGLPE 201

QY 226 DFIDSLEKTDYDDKYKTKLYPHYFPVWKKCCIPETRRMEMAFTNRCKENTIIQLQLLP 285

Db 202 DFLNSLEKMDGKLYTKLYPHYFPVWKKCCIPETRRMEMAFTNRCKENTIIQLQLLP 261

QY 286 LRPKVAKLLGYSTHADPVLMTAKSTSRVTAFLDDLKPLGEAREEFILNLKKEC 345

Db 262 LRAQKRLGLGFHFDYVLEMMNAKTSQVATFDELAKPLGEAREEFILNLKKEC 321

QY 346 KDRGFYDGGKINAWDLVYNTQTEELKYSIDQEFLEKEYFPYEVVTEGLNTYQELGLSLF 405

Db 322 ERGLFPDGRIRAWDMRYNNQVETRYCVDQNLKEYFPVQVTVHGLGIYQELGLSLF 381

QY 406 EQMTDAHVNKSVTLTYVKDKATGEVLGQFYLDLYPREGKYNHAAFCGLQGLLPDQSR 465

Db 382 HHEEGASAMHEDVRLYTARDAASGEVVGKPYLDLYPREGKYNHAAFCGLQGLLPDQSR 441

QY 466 MVAVALVNFVPSQVAGRPSLLRHDEVRYTFHEFGVHMQLCSQAEFAMFSGTHVEDFV 525

Db 442 QIAIAAWANFTKPTADAPSLQHDDEVETFFHEFGVHMQLCSQAEFAMFSGTHVEDFV 501

QY 526 EVPSQMLENNWVDVSLRLSKHYKDGSPDIADLLLEKLVASRLVNTGLTLRQIVLSKYD 585

Db 502 EAPSQMLENNWVDVSLRLSKHYKDGSPDIADLLLEKLVASRLVNTGLTLRQIVLSKYD 561

QY 586 QSLHTNTSLDAASEYAKYCEILGVAATPGTNMPATFGHLAGYDQYYGYLWSEVFSMD 645

Db 562 QALHTQTDADPAEYARLCQELILGVPATPGTNMPATFGHLAGYDQYYGYLWSEVFSMD 621

QY 646 MEYSCFKKGINPEVGMKYRNLLILKPGSLDGMMLHNLFRKREPQKAFMLSRGL 701

Db 622 MFHTRFQEGVLSKGMVMDYRSCILRPGGSEDASAMLRFLGRDPKQDAFLSKGL 677

RESULT 5

HYHUTH

N:Alternate names: endo-oligopeptidase A; endopeptidase 24.15; Pz-peptidase; soluble

C:Species: Rattus norvegicus (Norway rat)

C:Date: 30-Sep-1992 #sequence\_revision 17-Nov-1995 #text\_change 07-May-1999

C:Accession: S38760; A54152; S55999; A36165

R:McKie, N.; Dando, P.M.; Rawlings, N.D.; Barrett, A.J.

Biochem. J. 295, 57-60, 1993

A:Title: Thimet oligopeptidase: similarity to 'soluble angiotensin II-binding protein

A:Reference number: S38760; MUID:94029935

A:Accession: S38760

A:Molecule type: mRNA

A:Residues: 1-687 <MCK>

R:Pierotti, A.; Dong, K.W.; Glucksmann, M.J.; Orłowski, M.; Roberts, J.L.

Biochemistry 33, 622, 1994

A:Reference number: A54152; MUID:94114549

A:Contents: corrections





```
Db 180 KIKEIKKISVNSINFSKNLGEQEVITPTKQLEGVDPDSILTQETIKSDKSDNETLYK 239
QY 241 IYLYKPHYPVVKKCCIPETRRMEMAFNTRCKENTIILOQLLPLRTKVAKLGYSTHA 300
Db 240 VTFKYPIFPVVKMLASSAQKQAFADQNKV-PENEAIIITLKLRLDELASLLGYDTYA 298
QY 301 DFVLEMTAKSTSRVTAFTADLISQKLPKLGAEAEERFILNKKKECKDRGFEDGKINAWD 360
Db 299 NYNLYDKMAEDSTYWNFLNDLKDPLGRKELQVLQDMKAEDVKLNQGDADPNYYIWD 358
QY 361 LYVYMTQTEELKYSIDQELPKYFPIEVVTEGLLNTYQELGLSFEQMTDAH---VWN-- 415
Db 359 HRYYNDKYLLENFNVDLEKISEYFPLEATITGMLEYETFLNLFKFTETKDSQNSKSWHDD 418
QY 416 -KSVITYLYTKDKATGEVLQGFYLDLYPREGKYNHACFGQLGQCLLPDGRSMMAVAALVY 474
Db 419 VKOIAVWNMDPKSPNFGWITFDLHPRDKGYGHAANFGLSSFMTDDTTRSPYATVYC 478
QY 475 NFSQVAGRPSSLRRHDEVRTYFHEFGVHMHQICAQTDFAFSGT-NVETDFVEVPSQMLE 533
Db 479 NFSKSTKDKPSLLKNEIVTFPHELGCHGHDLVGQNKESRFGSPGVPWFVEAFPSQMLE 538
QY 534 NWVDVDSLRRLSKHYKDGSPADLLEKLVASRLVNTGLTLRQIVLSKYVDQSLHTNTS 593
Db 539 FWTWKNKELINLSHYKTEKIPESLINSIKTKHVNGALFTLRQLHFGLEDMKVHTCKD 598
QY 594 LDAAS-EYAKYCSSELLGAATPGTNMPA--TFGH-LAGGYDQYGYLYWSEVFSMDMEY 648
Db 599 LQNLISICDTFWNOLRDISLISNGGTLKGYDSFGHIMSDSYAGYGYLYWAEVATDMYH 658
QY 649 SCFKKEGINPVEGMYKRNLIILKPGSLDGMHNLHFLKREPQKAFLSRGL 701
Db 659 TFEAKDP-LNANGIOYRDIVLARGGLYDINDNLKEFLGREPSKDAFLKELGL 710

RESULT 7
C82353 oligopeptidase A VC0188 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: C82353
R:Heidelber, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
  chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B.
  l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
  Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: C82353
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-680 <HEI>
A:Cross-references: GB:AE004109; GB:AE003852; NID:g9654590; PIDN:AAF93364.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0188
A:Map position: 1
C:Superfamily: peptidyl-dipeptidase Dcp

Query Match 23.1%; Score 848; DB 2; Length 680;
Best Local Similarity 31.0%; Pred. No. 7.4e-44;
Matches 209; Conservative 128; Mismatches 304; Indels 34; Gaps 12;

QY 52 LSPEQIKRTEELIVQTKQYDVGMLGIEEYTYENCLOALADVVKYIVERTMLDFPOH 111
Db 17 IKPEHVKPAVEQAIDCRHTIDKV-LAENPQPSWESVIAPIEEVDRLSRKWSVSHMS 75
QY 112 VSSDREVAASFEADKRLSRFDIEMSRGDIERIVHQLQETCDLCKIKPEARYLEKSTK 171
Db 76 VVNSDELREAVESCLPLLSYEYSTWVGQHGKLFYAKTKESAEFAKLDRAQKQKSDSLR 135
QY 172 MGRNGLHLPEQVQVQNEIKSMKRMSELCDIFNKNLNEIDDTFLVFSKA-----ELGALPDD 226
```

```
Db 136 DFELSGLPLQEQRYGEISKRMSSELGSKFSNNVL--DATMGWTQKITDVNLLAGMPES 193
QY 227 FIDSLEKTDODDK-----YKITLKYPHYFPVVKKCCIPETRRMEMAFNTRC-----KE 274
Db 194 ALAAQAAAEAKGLEGYLLTLDIPSYLPMYTYCDNOLARKEYEAYVYTRASDRGPNAGKW 253
QY 275 ENTIILOQLLPLRTKVAKLGYSTHADFLVLEMTAKSTSRVTAFTADLISQKLPKLGAEAR 334
Db 254 DNSEIIAEOLKRLHEIARMLGFSTYSEKSLATKMAQTDOVLGFLNDLANKAKPQGEREV 313
QY 335 EFILNLKKKECKDRGFEDGKINAWDLYIYMTQTEELKYSIDQELPKYFPIEVVTEGLL 394
Db 314 EELRQFAESEFGVK-----QLELMDIAYYSEKQKHLFEISDEELRPYFPEQKVVGGLF 367
QY 395 NYIOELLGLSFPOMTDAHVNKSVLYITVKDKATGEVLQGFYLDLYPREGKYNHACFGL 454
Db 368 EYLSRLFGQVKEKRGQVDMHESVRFDFID-AQGTLRGSGFYLDLYAREHKKRGGAWMDEC 426
QY 455 QPGCLLPDGRSMMAVAALVWNFSQPVAGRPSSLRRHDEVRTYFHEFGVHMHQICAQTDFA 514
Db 427 RYKRTDTSALQTPVAYLTCNENRPVGDPALETHDEVTLTFHEFGHGHUMLTQVEVGA 486
QY 515 FSGTN-VETDFVEVPSQMLENWNVDLSRLSKHYKDGSPADLLEKLVASRLVNTGL 573
Db 487 VSGINGVPWDVELPSQFLENWCQEEALAFISGHYQTCGEPLPKAMLDKMLAAKNFQSAM 546
QY 574 LYLRLQIVLSKYVDQSLHTNTSLDAASEYAKYCSSEI-LGVAATPG---TNMPATFGLH-AGG 628
Db 547 FLRLQLEFLDFLTLLTYDPEVGPVKVLETLAEVKKVAVLPGLGLEWNRFSHSFSHIPAG 606
QY 629 YDGOYGYLYWSEVFSMDMEYSCFKKEGINPVEGMYKRNLIILKPGSLDGMHNLHFLKR 688
Db 607 YSAGYYSYLWAEVLSADAF-SRFEEGIPNRETGQSFLLNNILEMGGSEEPMELFKRFRGR 665
QY 689 EPNQKAFLSRGLHA 703
Db 666 EPQIDALLRHAGIAA 680

RESULT 8
S47718
Oligopeptidase A (EC 3.4.24.-) - Escherichia coli
N:Alternate names: prIC protein
C:Species: Escherichia coli
C:Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 29-Sep-1999
C:Accession: S47718; E65147; A43329
R:Plunkett, G.
  submitted to the EMBL Data Library, March 1994
A:Reference number: S47666
A:Accession: S47718
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-680 <PLD>
A:Cross-references: EMBL:U00039; NID:g46582; PIDN:AAB18474.1; PID:g466635
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;
  Riley, M.;
  Rose, D.J.; Mau, B.; Shaoh, Y.
  Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: E65147
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-860 <BLAT>
A:Cross-references: GB:AE000426; GB:U00096; NID:g1789910; PIDN:AAC76523.1; PID:g17899
A:Experimental source: strain K-12, substrain MG1655
R:Conlin, C.A.; Trun, N.J.; Silhavy, T.J.; Miller, C.G.
  J. Bacteriol. 174, 5881-5887, 1992
A:Title: Escherichia coli prIC encodes an endopeptidase and is homologous to the Salm
A:Reference number: A43329; MUID:92394891
A:Accession: A43329
A:Status: preliminary
A:Molecule type: DNA
```



C:Keywords: hydrolase; metalloproteinase

Query Match 22.2%; Score 814.5; DB 2; Length 681;  
Best Local Similarity 30.3%; Pred. No. 8e-42;  
Matches 206; Conservative 127; Mismatches 298; Indels 49; Gaps 15;

QY 52 LSPQIKTRTEELIVQTKQVDAVGMGLGTEETVYENCALQALADVEVKYIVERTMLDFPOH 111  
DB 19 IKPEHRAVEKLQDCRNTIQV--LKPHPTWENFILPTETNDRNLRAWSPVSHLNS 76

QY 112 VSSDEVRRASTEADKRLSRFDIEMSRGDIPIFIRVHLOETCDLQKIKPEARYLEKSTK 171  
DB 77 VKNSTELREAYOTCLPLLSSEYTWGQHGKLYNAYLAKNSAEFADYSTAQKKAATENSUR 136

QY 172 MGRKRLHLPQVQVQNEIKSMKRMSELCLDFNKNLNEEDTF---LVPSKAEGLGALPDF 227  
DB 137 DFELSGIGLSEKQQRQYGEIVARLSQNSQFNNV-LDAMGWELIENEAEALAGLPESA 195

QY 228 IDSLKTDODK----YKILKYPHYFPMKKKCCIPETRRRMEMAFNTRCKE-----E 275  
DB 196 LQAQQAFAKSLGKYRFTLEIPSLPVMTYCNALREEMYRAYATRASEOGPNAGKWD 255

QY 276 NYIILOQLPLTKVAKLIGYTHADVLEMTAKTSRVTAFLDDLQKLPGLGEARE 335  
DB 256 NSGVMEELTLRLVELAKLIGFNTYELSLATKMAENPQOVLDLHLAERAKPQGEKELQ 315

QY 336 FTLLNKKCKCKDRGEYDG--KINAWDLYYYMTQTEELKYSIDQELKXYFFIEVVTEGL 393  
DB 316 ELKGYCEKEF-----GVTELAPWDIGFSEKQKQHLAIINDELRYFPFENRVISGL 367

QY 394 LNTYQELGLSLPEQMTDAHVNMKSVTLTYVKDKATGEVLQGYLDLYPREGYNHA---A 450  
DB 368 FELIKRIFNIRAVRKGVDVHKVDFPDIDE--NDLQSGSYLDLYAREHKGAWMD 426

QY 451 CRGLQPGCLLPDGSRMMAVAALVNFSPQVAGRPSLLRHDEVRTYFHEFGHYMHQICAO 510  
DB 427 CJCRRKRL---DGSIEPTVAYLTCNFPNAPIGNKPAFLTHNEVTTLTFHEFGHHLMTQ 483

QY 511 DFARFSGTN-VETDFVEVPSQMLENNWVDVSLRLSKHYKDGSPADLLEKLVASRLV 569  
DB 484 DVSVDVAGINGVPWDVAVELPSPQFMENWCEEAALAFISGHYETGEPLPKELQLLAKNF 543

QY 570 NTGLLTLRQIVLSKYDQSLH-----NTSLDAASEYAKYCEILGVAATPGTNPATF 622  
DB 544 QAAMFLRQLEFGIFDFRLHHTFDKNTQLDLSKYSQVAVIKGV--DWARAPHSF 600

QY 623 GHL-AGGYDQGYGYLWSEVFSMDMFYSCFKKEGIMNPVGMKYNRLILKPGGSLDGM 681  
DB 601 SHIFAGGYAAGYSYLWAEVLSADA-YSRFEEGIFNPTTGSFLDEILTRGGSEPEMEL 659

QY 682 LHNELKRPNQKAFILMSRGL 701  
DB 660 FKRFRGEPQLDALLRHKGI 679

RESULT 11  
A42298  
thimet oligopeptidase (EC 3.4.24.15) - Salmonella typhimurium  
N:Alternate names: endo-oligopeptidase A  
C:Species: Salmonella typhimurium  
C:Date: 10-Jul-1992 #sequence\_revision 10-Jul-1992 #text\_change 29-Sep-1999  
C:Accession: A42298  
R:Conlin, C.A.; Miller, C.G.  
J. Bacteriol. 174, 1631-1640, 1992  
A:Title: Cloning and nucleotide sequence of opdA, the gene encoding oligopeptidase A in  
A:Reference number: A42298; MUID:92165739  
A:Accession: A42298  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-680 <CON>  
C:Cross-references: GB:M84574; NID:g154213; PIDN:AAA2172.1; PID:g154214  
C:Superfamily: peptidyl-dipeptidase Dcp

C:Keywords: hydrolase; metalloproteinase

Query Match 22.0%; Score 806; DB 2; Length 680;  
Best Local Similarity 30.4%; Pred. No. 2.6e-41;  
Matches 206; Conservative 129; Mismatches 300; Indels 42; Gaps 17;

QY 52 LSPQIKTRTEELIVQTKQVDAVGMGLGTEETVYENCALQALADVEVKYIVERTMLDFPOH 111  
DB 17 IKPEHVAVTKALADCAAEVGVVAHG-APYSWENLCOPLA--EADDVLGRIFSPI-SH 72

QY 112 VSSDK---EVRAASTADKRLSRFDIEMSRGDIPIFIRVHLOETCDLQKIKPEARYLEK 168  
DB 73 LNSVKNSPELREAYEQTLPLLSSEYTWGQHGKLYNAYRLRDGDHYATLNTAQKKAVDN 132

QY 169 SIKMKGRKRLHLPQVQVQNEIKSMKRMSELCLDFNKNLNEED--TFLVFSKAEGLALPD 225  
DB 133 ALURDFELSGIGLPEKQQRQYGEIATRLSELGNQYNNVLDATMGWTKLITDEAELAGMPE 192

QY 226 DFIDS----LEKTDGDKYKITLKYPHYFPMKKKCCIPETRRRMEMAFNTRC-----K 273  
DB 193 SALAAKAQAQAEKEGYYLLITDIPSYLPMVYCDNOALREEMYRAYSTRASDOGPNACK 252

QY 274 EENTILOQLPLTKVAKLIGYTHADVLEMTAKTSRVTAFLDDLQKLPGLGEARE 333  
DB 253 WNSPVMEEILLALRHQLAQLLGFENVAHESLATKMAENPQOVLDLTLAKRARPOGEKE 312

QY 334 REFILNKKCKCKDRGEYDGKINAWDLYYYMTQTEELKYSIDQELKXYFFIEVVTEGL 393  
DB 313 LAQLRAFAKAEF---GVE---ELQPMDIAYYSEKQKQHLYSISDEQLRYFPFENKAVNGL 366

QY 394 LNTYQELGLSLPEQMTDAHVNMKSVTLTYVKDKATGEVLQGYLDLYPREGYNHA---A 450  
DB 367 FVAVRIIGITAKERTDVDVHPEVRFELIDE--NNELRGSYLDLYAREHKGAWMD 425

QY 451 CRGLQPGCLLPDGSRMMAVAALVNFSPQVAGRPSLLRHDEVRTYFHEFGHYMHQICAO 510  
DB 426 CVGQMRKA---DGTLOKPVAYLTCNFPNVPNGKPAFLTHDEVITLTFHEFGHHLMTRI 482

QY 511 DFARFSG-TNVTDFVEVPSQMLENNWVDVSLRLSKHYKDGSPADLLEKLVASRLV 569  
DB 483 ETAGVSGISGVDPWDVAVELPSPQFMENWCEEAALAFISGHYETGEPLPKELDKMLAANY 542

QY 570 NTGLLTLRQIVLSKYDQSLHNTSLDAASEYAKYCEILGVAATPGT---NMPATFGHL 625  
DB 543 QAALFILRQLEFGIFDFRLHAEFNPQOGAKILETLEIKQVAVVPSPTWGRFPAFHSI 602

QY 626 -AGGYDQGYGYLWSEVFSMDMFYSCFKKEGIMNPVGMKYNRLILKPGGSLDGM 684  
DB 603 FAGGYAAGYSYLWADVLAADA-YSRFEEGIFNRETGQSFLLDNLTRGGSSEPEMELFKR 661

QY 685 FLKREPQKAFILMSRGL 701  
DB 662 FKRFRGEPQLDALLRHHGI 678

RESULT 12  
E83636  
oligopeptidase A PA0067 [imported] - Pseudomonas aeruginosa (strain PA01)  
N:Alternate names: endo-oligopeptidase A  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: E83636  
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A:Reference number: A82950; MUID:20437337  
A:Accession: E83636  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-681 <STO>  
A:Cross-references: GB:AE004446; GB:AE004091; NID:g9945886; PIDN:ANG03457.1; GSPDB:GN

A;Experimental source: strain PA01  
C;Genetics:  
A;Gene: prlc; PA0067  
C;Superfamily: peptidyl-dipeptidase Dcp

Query Match 21.7%; Score 794.5; DB 2; Length 681;  
Best Local Similarity 30.1%; Pred. No. 1.3e-40;  
Matches 202; Conservative 132; Mismatches 305; Indels 31; Gaps 12;

```
Qy 52 LSPQIKTRTEELIVQKQYDVGMLGIEBTVYENCLOALADVEKVIYVTRTMDLPPQH 111
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 18 IRPEHVPAIERILADNRAAIAARLETRQETQWKGLVLAENDRLRGAWSVPSHLNA 77
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 112 VSSDKEVRAASTEADKRLSRDIEMSRGDIIFERIVHLQETCDLKGKIPKERRVLEKSIK 171
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 78 VCSAEUREAEACLPESAYSTELGQNRLFEAYEALAKSPGAAGFDVAQKTLIEHALR 137
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 172 MGKRNGLHLPPQVQNEIKSMKRMSELICIDFNKML---NEDDTFLVFSKAEGLALPDDFI 228
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 138 DFLSGIDLPADKQKRYAEQVSRSELGSRFSNQLLDQATQAKHTVTDAAALAGLTSAS 197
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 229 DSEKTD---DYKTKLKYPPHYFPVMKKCCIPETRRRMEMAFNTRCKE-----EN 276
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 198 AQMKAAEAKGLDWLISPEPSYAYVYADRALREEVYAACTRASDQGNAGQNDN 257
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 277 TIILQQLPLRTKVAKLGYSTHADFVLEMTAKSTSRVTAFLDLSQKLKPLGEAREEF 336
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 258 GPVWEIIDLQELAGLLGFANYAELSLATKMAESSQVLSFLDLAVRSAPFAARDLEQ 317
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 337 ILNLKKCKDRGFYDGKINAMDLYYMTQTEELKSIDQEFLEKXFPPIEVVTEGLLNT 396
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 318 L---RAYAEOGCT---ELQSWDAGYVYAEKLEARYSVSQEALRAYFPVVKVLSGLFAI 370
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 397 YQELLGLSFEQMTDAHVNKSVTYLTVKDKATGVLGQFYLDLPRGKYNHACFGLOP 456
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 371 VERLYGQIREFHDFRMHADVRLEFLE---NGEHRGFYFDLYARANKRGGAWMGDARD 428
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 457 GCLLPDGRMAVAALVNFQVAGRPSSLRHDEVRTYFHEFGHVMHQICAOQDFARFS 516
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 429 RRRDAQGLIDPVAYLCNFTPAVNGRPALLTHDVTTLFHEFGHGLHLTLTRVEHAAAS 488
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 517 GTN-VETDFVEVPQOMLENVWVDLSRLRKHKKYKDGSPIDADDLLEKLVALSRVNTGLLT 575
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 489 GINGVAMDVELPQFMENWCWEPEGLALISAHYETGVVALPQDLLEKLAAKNFQSGMM 548
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 576 LRQIVLSKVDQSLTNTSLDAS--EYAKYCEILGVAAATPCTNNPA-TFGHL-AGGYDG 631
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 549 VRQLEFLDFELHA--THGDGRSVLQVLEGRDEAVAVMRPPAYNRFANSFAHIFAGYAA 607
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 632 QYGYLWSEVFSMDMFYSCFKKEGIMNPEVGMKYNLILKPGSLDGMDLNHLNFKREPN 691
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 608 GYYSYKWAELVSADAF-SRFEEGVFNPDGTGRAFEALARGSGRPMMLLPVDFRGREPS 666
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 692 QKAFMLSRGL 701
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 667 IDALLRHSGL 676
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 13  
B81224  
oligopeptidase A NMB0214 [imported] - Neisseria meningitidis (strain MC58 serogroup B)  
C;Species: Neisseria meningitidis  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C;Accession: B81224  
R;Rettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.  
Hickey, E.K.; Hart, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.  
Science 287, 1809-1815, 2000  
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve  
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A;Reference number: B81000; MUID:20175755  
A;Accession: B81224

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-678 <TET>  
A;Cross-references: GB:AE002379; GB:AE002098; NID:g7225435; PIDN:AAF40670.1; PID:g722  
A;Experimental source: serogroup B, strain MC58  
C;Genetics:  
A;Gene: NMB0214  
C;Superfamily: peptidyl-dipeptidase Dcp

Query Match 19.1%; Score 702; DB 2; Length 678;  
Best Local Similarity 28.4%; Pred. No. 5.2e-35;  
Matches 194; Conservative 126; Mismatches 302; Indels 60; Gaps 19;

```
Qy 55 EQIKTRTEELIVQ--KQVYDAVGMGLGTEEVY-YENCLOALADVEKVIYVTRT---ML 106
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 16 DQIKTEDIKPALQTAIAEAREQIAAIAKQATHTGWANTVEPLTG-----ITVRGRINGVV 70
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 107 DFPQHVSSDKVEVRAASTEADKRLSRDIEMSRGDIIFERIVHLQETCDLKGKIPKERRYL 166
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 71 SHLSVADTPELRNAVYNELMPEITVFTTEIGQDIETLYNREKTIKNSPEFDTLSPAQTKL 130
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 167 EKSIMKRNGLHLPEQVQNEIKSMKRMSELICIDFNKMLNEDDTFLV--SKAELGAL 223
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 131 NHDRLDFVLSGAELVPEQQAELAKLQTEGAQLSAKFSQNVLDATDAFGIYFDDAALAGI 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 224 PDD---FIDSLEKTDQDDKYKITLKYPPHYFPVMKKCCIPETRRRMEMAFNTRCKE---- 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 191 PEDALAMFAAAQSESKYGYKIGLQIHYLAVIQYADNRELREQIYRAYVTRASELSDDG 250
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 275 --ENTIILQQLPLRTKVAKLGYSTHADFVLEMTAKSTSRVTAFLDLSQKLKPLGEA 332
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 251 KFDNTANIDRTLNALQTLKLGFKNYAELSLATKMAQDTPEQVLNFLHDLARRAKPYAEK 310
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 333 E-----REFIINLKKCKDRGFYDGKINAMDLYYMTQTEELKYSIDQELKEVF 384
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 311 DLAEVKAFARE-SUNL-----ADLPQMDLGYASEKUREAKYAFSETEVKRYF 356
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 385 PIEVVTGELNTYQELLGLSPEQMTDAHVNKSVTYLTVKDKATGATGEVLGQFYLDLYPREG 444
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 357 PVGKVLNGLFAQIKKLGIGFTEKT-VPVHKDVRYPFELQ--NGETIGGYMDLYAREG 413
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 445 KYNHAAQFGLOPQCLLPDGRSMMAVAALVNFQVAGRPSSLRHDEVRTYFHEFGHVMH 504
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 414 KRGAWNDYKGRRRFSDGTQLQPLTAYLVCFAPVPVGGREARLSHDEILILFHETGHGLH 473
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 505 QICAQTDFAFESGTFN-VETDFVEVPQOMLENVWVDLSRLRKHKKYKDGSPIDADDLLEKL 563
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 474 HLLTQVDELGVSGINGVEMDAVELPQFMENFVWEYNVLAQMSAHEETGVPLPKELFDKM 533
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 564 VASRLVNTGLTLTRQIVLSKVDQSLTNTSLDASAYAKYCEI---LGVAAATPCTNNPA 620
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 534 LAKNFQGMFLVRQMEFALFDMMIYSEDDGRLKNMQQVLDVSRKKVAVIQPPEYNRFA 593
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 621 -TFGHL-AGGYDGQYGYLWSEVFSMDMFYSCFKKEGIMNPEVGMKYNLILKPGGSLDG 678
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 594 LSFGHIFAGGYAGYSYAWAEVLSADA-YAAFEESDDV-AATGKREWEILAVGGSRSA 651
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 679 MDMLNHLNFKREPNOKAFMLSRG 700
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 652 AESPKAFRGREPSIDALLRHSG 673
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 14  
G81996  
oligopeptidase A (EC 3.4.24.70) NMA0054 [imported] - Neisseria meningitidis (strain Z  
C;Species: Neisseria meningitidis  
C;Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C;Accession: G81996  
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo  
; Holroyd, S.; Jagsels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre  
Nature 404, 502-506, 2000  
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491

A:Reference number: A81775; MUID:20222556  
A:Accession: G81996  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-678 <PAR>  
A:Cross-references: GB:AL162752; GB:AL157157  
A:Experimental source: serogroup A, strain  
C:Genetics:  
A:Gene: prIC; NWA0054  
C:Superfamily: peptidyl-dipeptidase Dcp  
C:Keywords: hydrolase; metalloproteinase

Query Match	18.9%	Score 594;	DB 2;	Length 678;
Best Local Similarity	28.2%	Pred. No. 1.6e-34;		
Matches 190;	Conservative 131;	Mismatches 309;	Indels 44;	Gaps 18;

[illegible]

RESULT 15  
S76766  
hypothetical protein - *Synechocystis* sp. (strain PCC 6803)  
C:Species: *Synechocystis* sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_C  
C:Accession: S76766

R; Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yano, K. *Proc. Natl. Acad. Sci. USA* 95: 12285-12290 (1998).

D: A:Reference number: S74322; MUID:97061201  
 A:Accession: S76786  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-713 <KAN>  
 A:Cross-references: EMBL:D90916; GR:AB001339; NID:g1653715; PIDN:  
 A:Note: the nucleotide sequence was submitted to the EMBL Data L  
 C:Genetics:  
 A:Start codon: GTG  
 C:Superfamily: peptidyl-dipeptidase Dcp

Query Match	18.8%	Score 690.5	DB 2	Length 713
Best Local Similarity	27.9%	Pred. No. 2.8e-34		
Matches	196	Conservative 137	Mismatches 315	Indels 55
Gaps				
Qy	36	AMSSYTVAGRNVLWD--LSPEQIK----	TRTEELIVOTKOVYDAVGMGLGIEEVTYENCLQ	90
Db	24	AVANPLQGGDPLPFDKIQTDVDPVAVTTLVQELEAQLTELEKSV-----	QPTWEGLVE	77
Qy	91	ALADVEVKYIVERTMLDFPOHVSSDKKEVRAASTADKRLSRFDIEMSMRGDIFERIVHLQ	150	
Db	78	PLTAALEKLSWTGVTSHLMCMVKNPSPELRHGFQEQVQLPVGVGFTSRLGQSRPLYEAFVALR	137	
Qy	151	ETCOLGKIKPEARYLBKSIKMGKRNGLHLHPEQVQNEIKSMKXRMSELCIDFNKN--LNE	209	
Db	138	DSAEWGNLEPAQQRIVESNIREAELGCVGLTGKRDRAIQLAEALATKFSNNVLDA	197	
Qy	210	DTF--LVFSKAELGALPDDFTDSLEKTDGK-----	YKITLKYPHYFVPMKKCC	256
Db	198	QAFELKUTKEETIAGLPSSLALAAQTAQOQGEDNATPEEGPMLITLDYPSYLPFMKYSQ	257	
Qy	257	IPETRRMEAFNTRCKE---ENTIIQQLLPLRTKAKLLGYSTHADFVLENTAKSTS	313	
Db	258	REDLREKVYRAFTRRADGQWDNHPIERIALALEKAQLLGFTYAELSLRKMAPNVA	317	
Qy	314	RVTAFLDDLQSKLPLGEAREFETILNLKKCKEKGDFEYDGKINAWDLYYYMTQTEELKY	373	
Db	318	AVEKLLBEL--RVSPYKAAEQEF--ADLQK-----	FAGVEELHHWDTAYWSERQREQF	367
Qy	374	SIDOEFLKEYPPIEVVTGELLNTYQELLGLSFE--QMTDAHVNKNSVTLYTVKDKATGEVL	432	
Db	368	DFDAEALRPYFPLPQVLGEGFLALAKRULFGVTIOQASEEVPVMPDPQVQFFRVLDE--SGTAI	426	
Qy	433	GOFYLDLPREGKYNHAACFG--LQPGCLLPDG--SRMAVAALVNVFSPQVAGRPSLLR	488	
Db	427	ASFYLDAYSAPAEKRGGAWMADCLNRQELINGOKSLRPLVAYLICNOTPPVGDPQSLMT	486	
Qy	489	HDEVRTYFHEFGHYMHOICAQTODFARESG--TNVETDFEVPFSQMLENNWWDVDSLRLSK	547	
Db	487	FYEVTTLFHEFGHGLQLLTTVNYSGAAGINNVEWDAAVELPSPQFMENWCYDPTTLFSLAK	546	
Qy	548	HYKDGSPDIADLLEKIVASRLVNTGLTLTLQIVLSKYDQSLHTTSLDAASEYAKYCSEI	607	
Db	547	HYQGETLPQAEYEKILATRFNFTGSSAMLRQINPFLSDDLDELHURYRDPGPETIEQIGDRL	606	
Qy	608	LGVA-----TPTNMPATFGHL--AGGYDGOYGYGLWSEVFSMDMFYSCPKKEGTMN----	658	
Db	607	AKITTLPPPLPENAFCSFGHIFAGGYAAGYYSYKWAELVLSADAF--AAFEVGLDNEAAV	655	
Qy	659	PEVGMKYRNILKPGGSLDGMDLHNFLLKREPNOQAKFLMSRGL	701	
Db	666	OTTCKRRETIVLAGGSLAPAVVEKFERGPOPTAPLRRHGL	708	

Search completed: January 14, 2002, 18:57:49  
Job time: 798 sec

us-09-833-782-2.rpr

Tue Jan 15 09:01:18 2002

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 18:54:36 ; Search time 18.45 Seconds  
(without alignments)  
1371.775 Million cell updates/sec

Title: US-09-833-782-2  
Perfect score: 3668  
Sequence: 1 MIARCLLAVSLRRVGGSR.....FLKREPNQKAFILMSRGLHAP 704

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 170028 seqs, 35950645 residues

Total number of hits satisfying chosen parameters: 170028

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_New.\*  
1: /cgn2.6/ptodata/1/paa/PCT\_NEW\_COMB.pep.\*  
2: /cgn2.6/ptodata/1/paa/US06\_NEW\_COMB.pep.\*  
3: /cgn2.6/ptodata/1/paa/US07\_NEW\_COMB.pep.\*  
4: /cgn2.6/ptodata/1/paa/US08\_NEW\_COMB.pep.\*  
5: /cgn2.6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*  
6: /cgn2.6/ptodata/1/paa/US10\_NEW\_COMB.pep.\*  
7: /cgn2.6/ptodata/1/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	795	21.7	680	5	US-09-897-516-6500
2	732	20.0	655	5	US-09-897-516-6498
3	702	19.1	678	5	US-09-806-866-53
4	702	19.1	678	5	US-09-806-866-57
5	694	18.9	678	5	US-09-806-866-55
6	694	18.9	678	5	US-09-806-866-61
7	641.5	17.5	678	6	US-10-015-127-11537
8	550	15.0	691	5	US-09-897-516-8169
9	480.5	13.1	476	5	US-09-806-866-51
10	475	12.9	491	5	US-09-806-866-59
11	373	10.2	71	5	US-09-826-734-44
12	338.5	9.2	251	6	US-10-015-127-11338
13	130	3.5	1144	5	US-09-708-427-15046
14	130	3.5	1304	5	US-09-708-427-15045
15	130	3.5	1313	5	US-09-708-427-15044
16	122	3.3	1388	5	US-09-976-594-296
17	120	3.3	464	6	US-10-017-754-1906
18	117	3.2	1014	5	US-09-708-427-19883
19	117	3.2	1018	5	US-09-708-427-19882
20	117	3.2	1269	5	US-09-708-427-19881
21	116.5	3.2	701	5	US-09-708-427-3744
22	116.5	3.2	772	5	US-09-708-427-3743
23	115	3.1	1270	5	US-09-720-934-107
24	112.5	3.1	865	5	US-09-815-242-11536
25	111	3.0	961	5	US-09-708-427-12954
26	111	3.0	1027	5	US-09-708-427-12953

ALIGNMENTS

RESULT 1

US-09-897-516-6500  
; Sequence 6500, Application US/09897516  
; GENERAL INFORMATION:  
; APPLICANT: Corbin, David R.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Huesing, Joseph E.  
; APPLICANT: Krasomil-Osterfeld, Karina C.  
; APPLICANT: Malvar, Thomas M.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Spiridonov, Sergei  
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof  
; FILE REFERENCE: 38-21(51847)B  
; CURRENT APPLICATION NUMBER: US/09/897,516  
; CURRENT FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/215, 161  
; PRIOR FILING DATE: 2000-06-30  
; NUMBER OF SEQ ID NOS: 8409  
; SEQ ID NO 6500  
; LENGTH: 680  
; TYPE: PRT  
; ORGANISM: Xenorhabdus sp.  
US-09-897-516-6500

Query Match 21.7%; Score 795; DB 5; Length 680;  
Best Local Similarity 30.2%; Pred. No. 9e-47;  
Matches 207; Conservative 126; Mismatches 295; Indels 58; Gaps 17;

QY 51 DLSPEQIKTRTEELIVQTKOVYDAVGMGLTEETVYENCIALADVEKVIYIVERTMLDFQP 110  
DB 16 DIRPEHVPAVKETLANROVIEKI-LTENSFTWLSLCOPLSEQDKLSRMSPPVGHNLN 74  
QY 111 HVSSDKVEAASFEADKRLSRFDIENSMRGDIFERIVHQLQETCDLKIKKPEARYLEKSI 170  
DB 75 SVKNSPELREAYEQSLPLLESEFSTWLGQHKGLGYOAYKSESEVFNLSLQPKRSIEDTL 134  
QY 171 KMKRNGHLHPEQVQNEIKSMKRMSELCLDFNKNLNEEDTF----LVFSKAEALGALPDD 226  
DB 135 RDELSGIGLPEEKQRYGEGIAARMSELGSGQGNV-LDATMGWSKLITDENDLAGLPES 193  
QY 227 ----FIDSLEKTDQDKYKTKLKYPHYFPVYMKKCCIPETRRMEMAENTRC-----KE 274  
DB 194 AKAAKAAAEKQEGWLLTLNMPSYLPVNAVADNRELROEMTHAYTTRASDQGPAGKW 253  
QY 275 ENTIIQLQPLTKVAKLIGYSTHADFVLEMTAKSTSVTAFLDLDLSQKLPLGEAR 334  
DB 254 DNNAIAMELALRHELQQLLGFASYAEKSLATKMAKNPQOVLDFLNDLANRAHQCKKEL 313

Sequence 449, App  
Sequence 12952, A  
Sequence 512, App  
Sequence 11364, A  
Sequence 19314, A  
Sequence 19313, A  
Sequence 19312, A  
Sequence 515, App  
Sequence 11489, A  
Sequence 9203, Ap  
Sequence 9204, Ap  
Sequence 9203, Ap  
Sequence 21161, A  
Sequence 21160, A  
Sequence 21159, A  
Sequence 4897, Ap  
Sequence 2853, Ap  
Sequence 2854, Ap  
Sequence 2853, Ap

```
QY 335 EFILNKKCKDRGFEYD--KINAWDLYYYMTOTELKYSIDOEFLEKXFFPIEVVTEG 392
Db 314 E-----ELTDFARSHVGVKLEAWDTYYISEKQKHHSIDDEQLRPPPEORAVEG 365
QY 393 LUNTYQELGLGSFQMTDAHVNKSVTLTYVKDKATGEVLGYQLDLYPRECKYNHA--- 449
Db 366 LFEVVRRIYGITAKERNVDVETWHSVDVRFELYDE-KGELRGSFYLDLYAREHKRGAWMD 424
QY 450 ACFGLOPCGLLDGSRMAVAALVNFQVAGRPVAGRPSLLRHDEVRTYTFHEFGHVMHQICAQ 509
Db 425 DCAGRM---REFASGELQKPVAYLTFCNFKPVGDKPALETHDEVITFLFHEFGHGLHMLTE 481
QY 510 TDFARESGTN-VETDFEVPQSOMLENWVWDVDSLRLSKHYKDGSPDIADDLLEKLVASRL 568
Db 482 IETLDVAGLVGVPWDAVQSPQFQWENWCEPALAFISGHYETNEPLQPEMLDNMLAARN 541
QY 569 VNTGLTLTRQIVLSKVDOSLHTNTSLDAASEY--AK-----YCSEILGVAATPGTN 617
Db 542 YQSAFILLRQLEFGLDFRLH-----AEYDPAKGAQILPTLYSVKEQVSVVPSPEWN 593
QY 618 -MPATFGHL-AGGYDQYGYLWSVFSMDMFYSCFKKEGIMNPVGMKYNRLILKPGGS 675
Db 594 REPNSFSHIFNGGYAAGYSYLMADVLAADS-YSRFSEEGIFNRTTQGSFLDNILSRGGS 652
QY 676 LDGMDLHNLKREPKNQKAFLMRGL 701
Db 653 EDPWTLFTRFRGKPLDMLKLSAGI 678

RESULT 2
US-09-897-516-6498
; Sequence 6498, Application US/09897516
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)B
; CURRENT APPLICATION NUMBER: US/09/897,516
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215, 161
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 6498
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Xenorhabdus sp.
US-09-897-516-6498

Query Match 20.0%; Score 732; DB 5; Length 655;
Best Local Similarity 29.3%; Pred. No. 1.8e-42;
Matches 195; Conservative 123; Mismatches 308; Indels 40; Gaps 17;

QY 62 BELIVQTKQYDAVGMGLGIEVTVENCILQALADVEVKYIVERTMLDFPQHVSDDK---EV 118
Db 2 KEILAKYQVIEKV-VNENSQFTWDLNCLQPISEARNQF--ERVWSPV-SHLNSVKNTPEF 57
QY 119 RAASTEADKRLSRFDIENSMRGDIFERIVHLOETCDLGIKIPPEARLYLEKSIKWKGRNGL 178
Db 58 RVAYEQCLPMNSEFTWNGQKHGLYQAYKSLKSEFEFKLSQPKAVEDKLLDFTLSGI 117
QY 179 HLPQOVQNEIKSMKRMSELCDIDENKNLNEDD---TFVFSKAEGLALPDPDFIDSL--- 232
Db 118 ALPEKQKRYGEITARLSLSARFGNNVLDATMGWTKLVTDVCDLSGLPENIKSSAKAAA 177
QY 233 KTDDBK-YKITLKYPHYFPVMMKCCIPETRRMEMAFNTR-----CKEENTIILOOL 283
```

```
Db 178 KAKDLKGLMLTLDPPVYSAVMSYADDCCELREEVYTAWTRASDOGMNACKWDNNPIMEEI 237
QY 284 LPLRTKVAKLGLGYSTHADFPVLEMTAKSTSRVTAFLDLDLSQKLKPLGAEAREFINLAKK 343
Db 238 LALRHLEALLGLGKYNASLAKMTTKKEYLDVLNGLVDRAHGQGRMEELCEFAQS 297
QY 344 E-CKDRGFEYDGINAWDLYYYMTOTELKYSIDOEFLEKXFFPIEVVTEGLLNTYOELLG 402
Db 298 HCYVD-----ELESWDVRYSEKQKHHSIDDEQLRPPPEORAVEG 350
QY 403 LSFEQMTDAHVNKSVTLTYV-KDKATGEVLGYQLDLYPRECKYNHAACFGLQPCGLLP 461
Db 351 ITAKERNDVDTWNSDVRFEELYNDKE--ELLSGFYMDLYARAGKNGAWQSSYVDVDRMLA 408
QY 462 DGSRMMAVAALVNFQVAGRPVAGRPSLLRHDEVRTYTFHEFGHVMHQICAQTDFAFSGTN-V 520
Db 409 SGELQKPVAFINCFNFSFGPAGGKPALETHSEYSTLTFHEFGHGLQTLTTLTETLDVAGTNGV 468
QY 521 ETDFEVPQSOMLENWVWDVDSLRLSKHYKDGSPDIADDLLEKLVASRLVNTGLTLTRQIV 580
Db 469 PWDIVCEQSQIMENWCWEABALEFISGHYETNAPLPKEMLDNMLEAKNYQAAKMLRQLE 528
QY 581 LSKVDOSLHTNTSLDAASEYAKYCYSEIL-GVAATPGTN--MPATFGHL-AGGYDQYGY 635
Db 529 FGLDFRLHVEYDPEGAQILPILSVKEHVSVPSSDWERFPNSFHSIFNGGYAAGYS 588
QY 636 YLWSEVFSMDMFYSCFKKEGIMNPVGMKYNRLILKPGSLDGMMDLHNLKREPKNQKAF 695
Db 589 YMWADVLAADA-YSRFSEEGIFNRTTQGSFLDDYLGRRGSEDPLTLFTRFRGKPLDAL 647
QY 696 LMSRGL 701
Db 648 LKRLGI 653

RESULT 3
US-09-806-866-53
; Sequence 53, Application US/09806866
; GENERAL INFORMATION:
; APPLICANT: Frazer, Claire M.
; APPLICANT: Hickey, Erin
; APPLICANT: Peterson, Jeremy
; APPLICANT: Tettelin, Hervé
; APPLICANT: Venter, J. Craig
; APPLICANT: Maignani, Vega
; APPLICANT: Galeotti, Cesira
; APPLICANT: Mora, Manroza
; APPLICANT: Ratti, Giulio
; APPLICANT: Scarlato, Vincenzo
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagrazia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Neisseria Genomic Research
; FILE REFERENCE: CHIR-0313
; CURRENT APPLICATION NUMBER: US/09/806,866
; CURRENT FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: PCT/US00/05928
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: PCT/US99/25373
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: 60/103,794
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: 60/132,068
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-09-806-866-53
```



```

: TITLE OF INVENTION: Neisseria Genomic Research
:
: FILE REFERENCE: CHIR-0313
:
: CURRENT APPLICATION NUMBER: US/09/806,866
: CURRENT FILING DATE: 1999-10-08
:
: PRIOR APPLICATION NUMBER: PCT/US00/05928
: PRIOR FILING DATE: 2000-03-08
:
: PRIOR APPLICATION NUMBER: PCT/US99/25373
: PRIOR FILING DATE: 1999-10-08
:
: PRIOR APPLICATION NUMBER: 60/103,794
: PRIOR FILING DATE: 1998-10-09
:
: PRIOR APPLICATION NUMBER: 60/132,068
: PRIOR FILING DATE: 1999-04-30
:
: NUMBER OF SEQ ID NOS: 107
:
: SOFTWARE: PatentIn version 3.1
:
: SEQ ID NO 57
:
: LENGTH: 678
:
: TYPE: PRT
:
: ORGANISM: Neisseria meningitidis
:
: US-09-806-866-57

```

Query Match 19.1%; Score 702; DB 5; Length 678;

[illegible]

Query Match	19.1%	Score	702:	DB	5:	Length	678:
Best Local Similarity	28.7%	Pred.	No. 2.2e-40:				
Matches	194:	Conservative	127:	Mismatches	305:	Indels	50:
Gaps	18:						

  

QY	56	QIKRTRTEELIVQTKQVDAVGMGLGIEEVT-----YENCLOALADVEVKYIVERT-----ML	106
DB	17	QIQTEDIKPAVQT-AIAEARGOIAAQAHTGHWANTVERLTG-----ITERVGRINGVV	70
QY	107	DFPOHVSODKEVRAASTEADKRLSRFDIEMSRMGDIFERIVHLOQTCDLGIKPEARRYL	166
DB	71	SHLSNVVDTPELRAVYNELMPEITVFTTEIGQDIELYNRFKTIKNSPEFATLSPAQKTKL	130
QY	167	EKSIMKGRNCLHLPEQOVONEIKSMKRMSLCLDFNKN-LNEDDTFLVF--SKAELGAL	223
DB	131	DHDURDFVLSGAELVPERQAEALQTEGAQSAKFSONVLDATDAFYIFDDAAPTLAGI	190
QY	224	PDD-----FIDSLKTDODDKYKITLKYPHYFPMVKKCCIPETRRMEMAFNCRKE-----	274
DB	191	PEDALAMFAAAQSEGKYGKIGIOIPLYLAVIOYAGNRELPEQIYRAYVTRASELSNDG	250
QY	275	--ENTIILQOLLPLRTKVAKLLGYSTHADFVLEMTAKTSRVSATFADLDLSQKLKPLGEA	332
DB	251	KFDMTANIDRTLENALAKTLGFKFYAELSLATKMADTPQVLNFLHDLARRAKPYAEK	310
QY	333	EREFILMKKECKDRGEFYDGKN--AWDIYYMTQTTEELKYSIDQELKEYEPIEYVT	390
DB	311	D-----LAEVKAFAREHLGLADPOPDWLSYAGEKLREAKYAFSETEVKYPPVGVKVL	362
QY	391	EGLLNTYOELLGLSFEQMTDAHVANKSVTLTVKDKATGEVLQGYLDLYPRECKYNHAA	450
DB	363	AGLPAQIKKLKLGIGFAEKT-VPVWHKDVRYFELQO--NGKTIGGYMDLYAREKRGGAW	419
QY	451	CFGLOQCLPLPDRGMRMAALVYNFQSVAGRPSLLRHDEVRVYFFHFGHVMHQAICAT	510
DB	420	MNDYKGRRRRADGTLQLPTALVLCNFAPPVCGKKEARLSHDEILTLFHTGHLLHLLTQV	479
QY	511	DFARFSGTN-VETDFVEVPSQOMLWVWDVSLRRLSKHYKDGSPADDLLEKLVASRLV	569
DB	480	DELGVSGINGVEWDVAVELPSQFMENFVWEYVNLVAQMSAHEETGEBPLPKELFDKMLAAKNF	539
QY	570	NTGLLTLRQIVLSKVDOSLHTNTSLDAASEYAKYCSEI---LGVAAAPTGTNMPA-TFGHL	625
DB	540	QRGMFLYRQMEFALFDMMIYSESDCLKNQOQVLDVSRKEVAVIQPPEYRNFANSFGHI	599
QY	626	-AGGYDGOQCYGLMSEVSMDFYSCFKKEGIMPEVGMKNLILKPGSLDGMGLHNL	684
DB	600	FAGGSAGYSYSAWAEVLSTDA-YAAEESDDV-AATGKRFWQOEILAVGGSRAAESFKA	657
QY	685	FLKREPNOKAFMLSRG	700
DB	658	FRGREPSIDALLROSG	673

```

RESULT      4
US-09-806-866-57
: Sequence 57, Application US/09806866
: GENERAL INFORMATION:
: APPLICANT: Frazer, Claire M.
: APPLICANT: Hickey, Erin
: APPLICANT: Peterson, Jeremy
: APPLICANT: Tettelin, Herve
: APPLICANT: Venter, J. Craig
: APPLICANT: Masignani, Vega
: APPLICANT: Galeotti, Cesira
: APPLICANT: Mora, Manrosa
: APPLICANT: Ratti, Giulio
: APPLICANT: Scarselli, Maria
: APPLICANT: Scarlato, Vincenzo
: APPLICANT: Rappuoli, Rino
: APPLICANT: Pizza, Mariagrazia
: APPLICANT: Grandi, Guido

```

```
RESULT 5
US-09-806-866-55
; Sequence 55; Application US/09806866
; GENERAL INFORMATION:
; APPLICANT: Frazer, Claire M.
; APPLICANT: Hickey, Erin
; APPLICANT: Peterson, Jeremy
; APPLICANT: Tettelin, Herve
; APPLICANT: Venter, J. Craig
; APPLICANT: Masignani, Vega
; APPLICANT: Galeotti, Cesira
; APPLICANT: Ratti, Giulio
; APPLICANT: Scarselli, Maria
; APPLICANT: Scarlato, Vincenzo
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagratia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Neisseria Genomic Research
; FILE REFERENCE: CHIR-0313
; CURRENT APPLICATION NUMBER: US/09/806,866
; CURRENT FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: PCT/US00/05928
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: PCT/US99/25373
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: 60/103,794
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: 60/132,068
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-09-806-866-55

Query Match 18.98; Score 694; DB 5; Length 678;
Best Local Similarity 28.28; Pred. No. 7.7e-40;
Matches 190; Conservative 131; Mismatches 309; Indels 44; Gaps 18;

QY 55 EQIKTRTEELIVQT--KQYDVGMLGIEEVT-YENCLOALADVEKVIYVERT-----ML 106
DB 16 DQIKTEDIKPALQTAIAEAREQIAAIKQATHTGWANTVEPLTG-----ITVRGRIWGV 70
QY 107 DFPQHVSSDKVEVRAASTPADKRLSRFDIEMSMRGDIFPRIVHLOETCDLGKIKPEARYL 166
DB 71 SHLSNVTDTPELRAAYNELMPEITVFFTEIGQDIELYNRKFTIKNSPEFDTLSHAQTKL 130
QY 167 EKSIKMGKRNGLHLPQVQVONEIKSMKKRMSELCDIFNKN-LNEDDTFLVF--SKAELGAL 223
DB 131 NHDLRDFVLSGAELPPEQQAELAKLTQEGAQSAKFSQNLVDATDAFGIYFDDAAPLAGI 190
QY 224 PDD-----FIDSEKTDDBDKYITIKYHPFVPMKKCCIPETRRRMEMAFNRCKE----- 274
DB 191 PEDALAMFAAAQSECKGYKIGIQIPHYLAVIQYADNRKIREQIYRAYVTRASELSDG 250
QY 275 --ENTIILOQLPLRTKVAKLLGYSTHADFVLEMMNTAKSTSRVTAFLDLSOKLPLGEA 332
DB 251 KFDNTANDRTLENALQTAQLGLGFKNVAELSLSATKMDTPEQVLNLFHDLARRAKPYAEK 310
QY 333 ERFETLNKKKECKDRGEYDCKNANDLYYYMTQTTELKYSIDQEFLEKFFPEVTEV 392
DB 311 D--LAEYKAFARESLG---ADLPQWDVLGAGEKLEAKYAFSETEVKYPFPGVKVNLG 364
QY 393 LLNTYQELLGIFSEQMTDAHVWNKSVTLTYTVKDKATGVLGVQFVLDLYPRSGKYNHAACF 452
DB 365 LFAQIKKLYIGTFEKT--VPVHHKDVRYFEIQQ--NGETIGGVTDMLYARGGKGGAMWN 421
QY 453 GLQPGCLLPDGRMMVAALVYVNFSSQVAGRPSSLRHDEVRTPFEEFGVHMQICAQTFD 512
```

```
DB 422 DYKGRRRFSDGTLQLPTAYLVCNFTPPVGGKEARLSHDEILTLFHEHCHGHLHLLTQVDE 481
QY 513 ARFSGTN-VETDFVEVPSOMLENWVDVSLRRLSKYKDGSPADDDLEKLVASRLVNT 571
DB 482 LGVSGINGVDEWDAVELPSQFMENFVWVNVLAQMSAHEETGVPLPKELFDKMLAAKNFOR 541
QY 572 GLTLRLQIVLSKVDQSLHTNTSLDAASEYAKYCSFI--LGVAATPGTNMPA--TFGHL-A 626
DB 542 GMFLVRQMEFALFDMMIYSEDEGRKLNWQVLDVRKKEVAVVRPPEYRNFRANSFGHIFA 601
QY 627 GGYDQIYGYLWSEVFSMDPFYSCFKKEGINPEVGMKYRNLIILKPGSLDGMMLHNF 686
DB 602 GYGAGYYSYAWAEVLSDA--YAAFEESDDV--AATGKRFRWQEIILAVGSGRSAAESFKA 659
QY 687 KREPNOKAFMLSRG 700
DB 660 GREPSIDALLRHSG 673

RESULT 6
US-09-806-866-61
; Sequence 61; Application US/09806866
; GENERAL INFORMATION:
; APPLICANT: Frazer, Claire M.
; APPLICANT: Hickey, Erin
; APPLICANT: Peterson, Jeremy
; APPLICANT: Tettelin, Herve
; APPLICANT: Venter, J. Craig
; APPLICANT: Masignani, Vega
; APPLICANT: Galeotti, Cesira
; APPLICANT: Mora, Manroza
; APPLICANT: Ratti, Giulio
; APPLICANT: Scarselli, Maria
; APPLICANT: Scarlato, Vincenzo
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagratia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Neisseria Genomic Research
; FILE REFERENCE: CHIR-0313
; CURRENT APPLICATION NUMBER: US/09/806,866
; CURRENT FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: PCT/US00/05928
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: PCT/US99/25373
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: 60/103,794
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: 60/132,068
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-09-806-866-61

Query Match 18.98; Score 694; DB 5; Length 678;
Best Local Similarity 28.28; Pred. No. 7.7e-40;
Matches 190; Conservative 131; Mismatches 309; Indels 44; Gaps 18;

QY 55 EQIKTRTEELIVQT--KQYDVGMLGIEEVT-YENCLOALADVEKVIYVERT-----ML 106
DB 16 DQIKTEDIKPALQTAIAEAREQIAAIKQATHTGWANTVEPLTG-----ITVRGRIWGV 70
QY 107 DFPQHVSSDKVEVRAASTPADKRLSRFDIEMSMRGDIFPRIVHLOETCDLGKIKPEARYL 166
DB 71 SHLSNVTDTPELRAAYNELMPEITVFFTEIGQDIELYNRKFTIKNSPEFDTLSHAQTKL 130
QY 167 EKSIKMGKRNGLHLPQVQVONEIKSMKKRMSELCDIFNKN-LNEDDTFLVF--SKAELGAL 223
```

```
Db 131 NHDRLRDLVSLGAEPLPEQQAELAKLQTEGAQLSAKFSQNVLDATDAFGIYFDAAAPLAGI 190
QY 224 PDD----FIDSLEKTDYKXITLKYPHYFPYMKKCCIPETRRMEMAENTRCKE----- 274
Db 191 PEDALAMFAAAQSEKGTGKIGLOIPHYLAVIQYADNRKLRQIYRAXVTRASELSDDG 250
QY 275 --ENTIILOQLPLRTRKAKLLGYSTHADFVLEMTAKSTSRVTAFLDLDLSOKLKLPLGEA 332
Db 251 KPDNTANIDRTLENALQAKLLGFKNYAELSLATKWADTPEQVLFNLHDLARRAKPYAEK 310
QY 333 ERFILNLKKCKDRGFYDGGKINAWDLYYYMTQTEELKYSIDQEFLEKYPPIEVVTEG 392
Db 311 D---LAEVRAFARESLGL---ADLPQWDLGYAGEKLRKAYAFSETEVRKYPVGVKVLNG 364
QY 393 LLNTYOELLCLSPQMTDAHVNKSVTLYTVKDKATCEVLGQFYLDLYPREGKYNHAACF 452
Db 365 LFAQIKKLYGIGTEKT--FVWHKDVRYFELQO--NGETIGGYMDLYAREGKRGGAWN 421
QY 453 GLQPGCLLPDGSRMMAVAALVNFSPVAGRPSLLRHDEVRTYFHEFGHVMHOICAOTDF 512
Db 422 DYKGRFRSDGTQLTALYLCNFTPPVGKKEARLSHDELLTLFHTGHLHLLTQVDE 481
QY 513 ARSGTN-VETDFEVPQSOMLENWVDLSRLSKHYKDGSPIDADDLLEKLVASRLVNT 571
Db 482 LGVSGINGVEMDAVELPSQFMENFVWYNLAQMSAHEETGVPLPKELFDKMLAAKNFOR 541
QY 572 GLTLRQIVLSKVDQSLHTNTSLDAASEYAKYCSEI---LGVNATPGTNMPA--TFGHL-A 626
Db 542 GMFLVRQMEFALFDMYIYSEDDEGRLLKNQOVLDSVRKEVAVVVRPPYRNFRANSFGHIFA 601
QY 627 GYDGOYGYGLWSEVPSDMFYSCFKKEGIMNPEVGMKYRNLLIKPGSLDGMDLHNLFL 686
Db 602 GGSAGYYSYANAEVLSADA-YAAFEESDDV-AATGKRFWQEIILAVGGRSRAESKAFR 659
QY 687 KREPNOKAFILMSRG 700
Db 660 GREPSIDALLRHSG 673

RESULT 7
US-10-015-127-11537
; Sequence 11537, Application US/10015127
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Sphingomonas elodea genome sequences and uses thereof
; FILE REFERENCE: 38-10(15806)B
; CURRENT APPLICATION NUMBER: US/10/015,127
; PRIOR FILING DATE: 2001-10-29
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 14357
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Sphingomonas elodea
US-10-015-127-11537

Query Match 17.5%; Score 641.5; DB 6; Length 678;
Best Local Similarity 28.1%; Pred. No. 3,1e-36;
Matches 189; Conservative 122; Mismatches 327; Indels 35; Gaps 16;

QY 52 LSPEQIKTTEELIVQTKQYDVGMLGIEEYVYENCALADVEVKYIVERTMLDFFQH 111
Db 17 IHPDQIAPALDAVIAEHAQAAVERV--IASESRFDDVWMPLEAETAIDALWSTVSHLRG 74
QY 112 VSSDKVEAASAEADKRLSRFDIEMSRGDIPIRIVHLQETCDLGLKIKPEA--RYLEKSI 170
Db 75 VADTPELRAAHAGAGQARLIENQLAVNQALAEVLVALTATPDFA--ARPOADRAAYEHMV 133
QY 171 KMKGRNGLHLPQEQVQNEIKSMKRMSELCTDFNKNLNEDDTFLVFSKAE-----LGLALPD 225
```

```
Db 134 RDRFLSGVALDAEAKARFAAASVSELSTLSTGEG--NAVLDATDAWPEHIEDPALLAGISPA 192
QY 226 D---FIDSLEKTDYKXITLKYPHYFPYMKKCCIPETRRMEMAENTRCKE----- 274
Db 193 DMAIFADAARAKLEGWVYVTLQAPSAYVLTFAENRDLRARLYRAYGTASDQGPAGHAGOF 252
QY 275 ENTIILOQLPLRTRKAKLLGYSTHADFVLEMTAKSTSRVTAFLDLDLSOKLKLPLGEAER 334
Db 253 DNGPRIAALLELRREAAQLGPPDPVANSLATKMPAAGEVLAFLRDLGRRARPAAQAE- 311
QY 335 EFLINLKKCKDRGFYDGGKINAWDLYYYMTQTEELKYSIDQEFLEKYPPIEVVTEGL 394
Db 312 --FAELSAFAARELGID---TLQPMWDVFPASDRLARAYAVDEQEVRAHFPPVERVIAGWQ 366
QY 395 NYIOELLGLSFQMTDAHVNKSVTLYTVKDKATCEVLGQFYLDLYPREGKYNHAAACFGL 454
Db 367 ALLARLFGVRLVARPDVAVYHPDACYDYVD--AEGVVIAGVYTDLHARAGKRSAGWMAQA 425
QY 455 QPGCLLPDGS--RMMAVAALVNFSPVAGRPSLLRHDEVRTYFHEFGHVMHOICAOTDFA 513
Db 426 RP--RUNDGNVRPVPAYLYVCNFPKTEGSPSLSHPEVVVTLHETGCHLHLFTVRDRP 483
QY 514 RFSGTN-VETDFEVPQSOMLENWVDLSRLSKHYKDGSPIDADDLLEKLVASRLVNTG 572
Db 484 NIAGTSGFEMDAVELPSQLMEDFAWDROVLRGMSGHATGAPLPDVLFDKLIARREFLAG 543
QY 573 LTLRQIVLSKVDQSLHTNTSLDAASEYAKYCSEIILGVNATPGTN--MPATFGHL-AGGYD 630
Db 544 MALVRQIEFALFDLLHLGTMGSDPMKVIKAVRDEVAVVVRPPPEWHFRPHAFSHIFAGGYA 603
QY 631 GOYGYGLWSEVPSDMFYSCFKKEGIMNPEVGMKYRNLLIKPGSLDGMDLHNLFLKREP 690
Db 604 SGYISYLMAEVLAADGFGQ--FAEAGLIDRATDRFDEVLRGASRPAAESFRAFGRDP 662
QY 691 NOKAFLMSRGLHA 703
Db 663 DPQAMLLRHGLTA 675

RESULT 8
US-09-897-516-8169
; Sequence 8169, Application US/09897516
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)B
; CURRENT APPLICATION NUMBER: US/09/897,516
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215, 161
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 8169
; LENGTH: 691
; TYPE: PRT
; ORGANISM: Xenorhabdus sp.
US-09-897-516-8169

Query Match 15.0%; Score 550; DB 5; Length 691;
Best Local Similarity 26.2%; Pred. No. 6,1e-30;
Matches 163; Conservative 117; Mismatches 267; Indels 74; Gaps 16;

QY 122 STEADKRLSRFDIEMSRGDIPIRIVHLQETCDLGLKIKPEARKYLEKSIKMGKRNGLHLP 181
Db 100 SEEMSPKLAAMNDEIMLNSKLFNRKTIYQOQRETNLDSLSRLVEVIYKQFELAGANLS 159
```



Db 227 FSDGTLQPLPAYLVCNFPAPVGGREARLSDHEILILFHETGHLHLHTQVDELGVSGIN 286  
QY 520 -VETDFVEVPSQMLENNWVDVDSIRLSLKHVKDGSPIADDDLEKLVASRLVNTGLTLRQ 578  
Db 287 GVXWDVPELSPQMFVWYINVLAXSAHEETGVPLKELDKXLAANKNFOXGMFXVRQ 346  
QY 579 IVLSKYDQSLHTNTSLDAASEYAKYCSSEI---LGVAAATPGTNMPA-TFGHL-AGGYDGGY 633  
Db 347 XEPALFDMMIYSEDDCRLKNWQVLDVSRKKVAVIOPPEYNRFALSFCHIFAGGYSAAX 406  
QY 634 YGLWSEVFMDFSCFKKEGTMNPEVGMKYNRLILKPGGSLDGMMDLHNFILKREPNOK 693  
Db 407 YSYAWAEVLSADA-YAAFEESDDV-AATGKRFEQTEILAVGXSRGSAESKAFKRGREPSID 464  
QY 694 AFLMSRG 700  
Db 465 ALLRHSG 471

RESULT 10  
US-09-806-866-59  
; Sequence 59, Application US/09806866  
; GENERAL INFORMATION:  
; APPLICANT: Frazer, Claire M.  
; APPLICANT: Hickey, Erin  
; APPLICANT: Peterson, Jeremy  
; APPLICANT: Tettelin, Hervé  
; APPLICANT: Venter, J. Craig  
; APPLICANT: Masignani, Vega  
; APPLICANT: Galeotti, Cesira  
; APPLICANT: Mora, Manroza  
; APPLICANT: Ratti, Giulio  
; APPLICANT: Scarselli, Maria  
; APPLICANT: Scarlato, Vincenzo  
; APPLICANT: Rappuoli, Rino  
; APPLICANT: Pizza, Mariagrazia  
; APPLICANT: Grandi, Guido  
; TITLE OF INVENTION: Neisseria Genomic Research  
; FILE REFERENCE: CHIR-0313  
; CURRENT APPLICATION NUMBER: US/09/806,866  
; PRIOR FILING DATE: 1999-10-08  
; PRIOR APPLICATION NUMBER: PCT/US00/05928  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: PCT/US99/25373  
; PRIOR FILING DATE: 1999-10-08  
; PRIOR APPLICATION NUMBER: 60/103,794  
; PRIOR FILING DATE: 1998-10-09  
; PRIOR APPLICATION NUMBER: 60/132,068  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 59  
; LENGTH: 491  
; TYPE: PRT  
; ORGANISM: Neisseria gonorrhoeae  
US-09-806-866-59

Query Match 12.9%; Score 475; DB 5; Length 491;  
Best Local Similarity 28.0%; Pred. No. 5.4e-25;  
Matches 137; Conservative 88; Mismatches 222; Indels 42; Gaps 12;  
QY 56 QIKTRTEELIVQTKYVDVAGMLGIEVET---YENCLOALADVEVKYIVERT-----ML 106  
Db 17 QIKTEDIKPAVQT-AIAEARGQAAVKAQTHGTWANTVERLTG-----ITERVGRWGVV 70  
QY 107 DFPQHVSSKVEVRAAATEADKRLSDIEMSRGDIFFERIVHLQETCDLGKIKPEARRYL 166  
Db 71 SHLNSVVDPELRAVYNELMPETVFFTEIGQDIELYNRFKTIKNSPEATLSPAKTKL 130  
QY 167 EKSINKGRNGLHLPEQVONEIKSMKKRMSELICIDFNKN-LNEDDFFLVF--SKAELGAL 223  
Db 131 DHDLRDFVLSGAELEPPERQAEALAKLQTEGAQLSAKFSQNVLDATDAFGIYFDDAAPLAGI 190

QY 224 PDD----FIDSLKTDKDKYKITLKYPHYPPVMKCCIPETRRMEMAFNTRCKE----- 274  
Db 191 PEDALAWFAAAOSEKGTGKIGLIQIPIHVLAVIQYAGNRELREQIYRAYVTRASELSNDG 250  
QY 275 --ENTTILOQLPLRTKVAKLGLGYSTHADFLVEMNTAKSTSRVTAFLDLDLSOKLPIGEA 332  
Db 251 KFDNTANIDRTLENALKTAKLGLGFKNYAELS LATKNAADTPEQVVLNFDLARRAKPYAEK 310  
QY 333 EREFIILNKKKCKDRGFEDYDGIN--AMDLYYMTQTEELKYSIDOEFLKEYFPPIEVVT 390  
Db 311 D-----LAEYKAFAREHGLADPQPDLSYAGEKLEAKYAFSETEVKKYFPYGVKVL 362  
QY 391 EGLNTYQELGLSEFQMTDAHVWNKSVTLTYTKDKATGEVLGQFYLDLYPREGKYNHAA 450  
Db 363 AGLFAQIKKIDYIGIFAECT--VPVWHKDRVYFELQQ--NGTKIGVYMDLYFARGKRGGAW 419  
QY 451 CFGLOPGCLLPDGRMMAVAALVNVNFSQPVAGRPSSLRHDEVRTYFHEFGHVMHOICAQT 510  
Db 420 MNDYKGRRRFADGTQLQPLTAYLVCNFAFPVGGKEARLSHDEILTLFHETGHLHLHTQV 479  
QY 511 DPARFSGTN 519  
Db 480 DELGVSGIN 488

RESULT 11  
US-09-826-734-44  
; Sequence 44, Application US/09826734  
; GENERAL INFORMATION:  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Mishra, Vishnu S.  
; APPLICANT: Leach, Martin D.  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Zerhusen, Bryan D.  
; APPLICANT: Kekuda, Ramesha  
; TITLE OF INVENTION: Novel Polynucleotides and Polypeptides Encoded Thereby  
; FILE REFERENCE: 15966-754  
; CURRENT APPLICATION NUMBER: US/09/826,734  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: 60/195,576  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 270  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 44  
; LENGTH: 71  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-09-826-734-44  
Query Match 10.2%; Score 373; DB 5; Length 71;  
Best Local Similarity 100.0%; Pred. No. 3.7e-19;  
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 182 EQVONEIKSMKKRMSELICIDFNKNLNEDDTFLVFSKAEALGALPDDFIDSLEKTDKDKYKI 241  
Db 1 EQVONEIKSMKKRMSELICIDFNKNLNEDDTFLVFSKAEALGALPDDFIDSLEKTDKDKYKI 60  
QY 242 TLKYPHYFPVM 252  
Db 61 TLKYPHYFPVM 71  
RESULT 12  
US-10-015-127-11338  
; Sequence 11338, Application US/10015127  
; GENERAL INFORMATION:  
; APPLICANT: Bower, Stanley G.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; TITLE OF INVENTION: Sphingomonas elodea genome sequences and uses thereof

FILE REFERENCE: 38-10(15806)B  
CURRENT APPLICATION NUMBER: US/10/015.127  
CURRENT FILING DATE: 2001-10-29  
PRIOR APPLICATION NUMBER: US 60/252,455  
PRIOR FILING DATE: 2000-11-22  
NUMBER OF SEQ ID NOS: 14357  
SEQ ID NO 11338  
LENGTH: 251  
TYPE: PRT  
ORGANISM: Sphingomonas elodea  
US-10-015-127-11338

Query Match 9.2%; Score 338.5; DB 6; Length 251;  
Best Local Similarity 31.9%; Pred. No. 5e-16;  
Matches 75; Conservative 45; Mismatches 108; Indels 7; Gaps 3;  
QY 475 NPSQVAGRPRLRLHDEVRTTFEFGHVMHQICATDFARFSGTNVETDFEVPQSMLEN 534  
DB 7 NPTKPAQGPALISDDVTFEFGHGLHGLFANQTPSVSGTNVARDFAEFPQFNEH 66  
QY 535 WYWDVDSLRRLSKHYKDGSPADDDLEKLVASRLVNTGLTLRLQIVLSKVQDSLHTNTSL 594  
DB 67 WALDPKVLPHYAVNTKOGSVIPQALVDKIKRAGTNSGYSGEALAAEMDMWSHSLAA 126  
QY 595 DAASEYAKYCEILGVAATPGTNMP-----ATFGHLAG-GYDGOYGYLWSEVFSDMFY 648  
DB 127 DGKQADAFEAQKALAATGLDVTDPYPRYSSVFLHWGNGYSAGYYAWTKMLSANAF- 185  
QY 649 SCFKKEGTMNPEVMGMYNRLILKPGSGLDGMDMLNHLKRPNQKAFMLSGRLHA 703  
DB 186 NWFQHGGMTRANGORFREMVLKSGHTEDYAPMFRFNGADFPQVAPLKLKDLGLNA 240

RESULT 13  
US-09-708-427-15046  
; Sequence 15046, Application US/09708427  
; GENERAL INFORMATION:  
; APPLICANT: N. ALEXANDROV et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1243P  
; CURRENT APPLICATION NUMBER: US/09/708.427  
; CURRENT FILING DATE: 2000-11-09  
; NUMBER OF SEQ ID NOS: 85364  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 15046  
; LENGTH: 1144  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: 1..1144  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc.feature  
; LOCATION: 1..1144  
; OTHER INFORMATION: Ceres Seq. ID 1828629  
US-09-708-427-15046

Query Match 3.5%; Score 130; DB 5; Length 1144;  
Best Local Similarity 18.2%; Pred. No. 0.86;  
Matches 111; Conservative 84; Mismatches 194; Indels 220; Gaps 24;  
QY 60 RTEELIVOTKQYDAVGML-----GIEEVTYENCLOALADVEVKYI 100  
DB 520 KIEELSVANESLADNVDTLQISVQESKDKEREVAYLKKIEELSVAN--ESLVKQETKL- 576  
QY 101 VERTMLDPFQVSSD-KEVRAASTEADKRLSRFDIEMSRGDIFFERIVHLQETCDLGR-I 158  
DB 577 -----QHIDQAEELRGREASHLKKIEELSK-----NENLVDNVANMQNIAESKDL 624  
QY 159 KPEARLYLEKSIKMGKRNGLHLPEQVQVQNEIKSMKRMSELCDIFNKNLNEDDTFLVFSKA 218

DB 625 REREVAYLKKIDELSTANGT-LADNVTN-----LQNISE-----ENKELRERETILLKKA 674  
QY 219 ELGALPDDFIDSLEKTDODDKYKITLKYPHYFPVMKKCCIPETRRRMENAFNTRCKEENTI 278  
DB 675 ELSNELSLVDKASKL-----QTVVQE-----NEELRERETA 706  
QY 279 ILQQLPLRTKVAKLLGYSTHADFVLEMTAKTSRVTAFDLDLSQKLPGLGEAREFILL 338  
DB 707 YLKKI-----EELSKLHEILSDQETKLQ-----IS 731  
QY 339 NLKKECKDRGFEYDKINANDLYYYTQTTEELKYSIDQEFLEKVFPIEVTEGLLNTYQ 398  
DB 732 NHEKEELKER-----ETAYLKKIEELSKYQEDLLNKEN 764  
QY 399 ELLGLSEF-----OMTDAHVMNKSVTLYTVKDKATGEVLGQGVFLDLYPREGYNHAAFCG- 453  
DB 765 ELHGMVVEIDLRKSDSLAOKKI-----EELSNNASLLIKENELQAVVCENE 812  
QY 454 -----LQPGCLLPDGSRRMVAALVNFVFSQPVAGRP 484  
DB 813 ELKSKQVSTLKTIDELSDLKQSLIHKEKELQ--AAIVENEKLEKAAALSQRIEELTNLK 870  
QY 485 SLL--RHDEVITYEHE-----FGHVMHQICATDFARFSGTNVETDFVEVP 528  
DB 871 QTLIDKQNELQGVFHENEELKAKEASSLKKIDELH-----LEQSWLEKESFQRTV 922  
QY 529 SOMLENWYWDVDSLRRLSKHYKDGSPADDDLEKLVASRLVNTGLTLRLQIVLSKVDQ-S 587  
DB 923 QENLELKTQDALAAKKI-----EELSKLKESELLEK-----ETELKCREAAALEKMEEPS 971  
QY 588 LHTNTSLDA 596  
DB 972 KHGNSLNS 980

RESULT 14  
US-09-708-427-15045  
; Sequence 15045, Application US/09708427  
; GENERAL INFORMATION:  
; APPLICANT: N. ALEXANDROV et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID  
; FILE REFERENCE: 2750-1243P  
; CURRENT APPLICATION NUMBER: US/09/708.427  
; CURRENT FILING DATE: 2000-11-09  
; NUMBER OF SEQ ID NOS: 85364  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 15045  
; LENGTH: 1304  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: 1..1304  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc.feature  
; LOCATION: 1..1304  
; OTHER INFORMATION: Ceres Seq. ID 1828628  
US-09-708-427-15045

Query Match 3.5%; Score 130; DB 5; Length 1304;  
Best Local Similarity 18.2%; Pred. No. 1;  
Matches 111; Conservative 84; Mismatches 194; Indels 220; Gaps 24;  
QY 60 RTEELIVOTKQYDAVGML-----GIEEVTYENCLOALADVEVKYI 100  
DB 680 KIEELSVANESLADNVDTLQISVQESKDKEREVAYLKKIEELSVAN--ESLVKQETKL- 736  
QY 101 VERTMLDPFQVSSD-KEVRAASTEADKRLSRFDIEMSRGDIFFERIVHLQETCDLGR-I 158  
DB 737 -----QHIDQAEELRGREASHLKKIEELSK-----NENLVDNVANMQNIAESKDL 784







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: January 14, 2002, 18:57:51 ; Search time 1811.41 Seconds  
(without alignments)  
12546.757 Million cell updates/sec

Title: US-09-833-782-1

Perfect score: 2115

Sequence: 1 atgatgcgcgcgtgcctttt.....gaggcctgcgtcgtga 2115

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estfun:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estom:\*

5: em\_estpl:\*

6: em\_estba:\*

7: em\_estro:\*

8: em\_estov:\*

9: em\_hic:\*

10: gb\_estl:\*

11: gb\_est2:\*

12: gb\_hic:\*

13: gb\_gss:\*

14: em\_gss\_fun:\*

15: em\_gss\_hum:\*

16: em\_gss\_inv:\*

17: em\_gss\_pln:\*

18: em\_gss\_pro:\*

19: em\_gss\_rod:\*

20: em\_gss\_vrt:\*

21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	622	29.4	625	10	AI952756
2	611	28.9	789	11	BI335570
3	570	27.0	579	11	BI335570
4	506.8	24.0	902	11	BI335570
5	479.4	22.7	745	11	BI335570
6	475.6	22.5	619	11	BI335570
7	472.6	22.3	745	11	BI335570
8	455.6	21.5	559	10	BE311015
9	431.4	20.4	509	10	AW786845
10	419.8	19.8	509	10	AW732454
11	416.8	19.7	511	10	AI386253
C 12	390.6	18.5	757	10	AW003054

C 13	376.8	17.8	680	11	BG380774
14	373.2	17.6	675	11	BF719641
15	358.2	16.9	417	10	AA152501
C 16	358	16.9	680	10	BE630187
17	344.8	16.3	447	10	AA066229
C 18	341.4	16.1	750	10	AI326417
19	331.4	15.7	366	10	AW859092
20	331.4	15.7	889	11	BG245790
21	328.8	15.5	560	10	BE627164
22	324.6	15.3	654	11	BF535533
23	318.8	15.1	984	11	BF974442
24	315.2	14.9	660	11	BI393320
25	306.4	14.5	759	11	BG967757
26	302.4	14.3	517	10	AI195498
27	299.4	14.2	818	10	AL521398
28	292.8	13.8	857	11	BF758404
29	289	13.7	765	10	AI322656
30	288.8	13.7	956	10	AL553474
31	280	13.2	776	11	BI327686
32	278.8	13.2	777	11	BI327686
C 33	272.2	12.9	462	10	AI195785
34	265.8	12.6	902	11	BF664439
35	264	12.5	717	11	BF528020
36	261.6	12.4	907	11	BE875429
37	258.6	12.2	901	10	AL551921
38	258.2	12.2	729	11	BG260342
39	257.2	12.2	725	11	BI194940
C 40	256.8	12.1	382	10	AW742681
C 41	255	12.1	706	11	BF222737
42	252.2	11.9	759	11	BG686225
43	252.2	11.9	530	11	BI314174
44	251.2	11.9	949	11	BI180024
45	250.6	11.8	748	11	BG972692

#### ALIGNMENTS

RESULT 1	AI952756	625 bp	mrna	EST	08-MAR-2000
AI952756/c	wr50b06.x1	NCI-CGAP_Ut1	Homo sapiens	cdna	IMAGE:2491091 3'
LOCUS	similar to SW:NEUL_PIG	Q02038	NEUROLYSIN	PRECURSOR ;	mrna
DEFINITION	sequence.				
ACCESSION	AI952756				
VERSION	AI952756.1	GI:5745066			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.				
TITLE	1 (bases 1 to 625)				
JOURNAL	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				
	Unpublished (1997)				
	Contact: Robert Strausberg, Ph.D.				
	Email: cgapbs-remail.nih.gov				
	Tissue Procurement: Christoph Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.				
	CDNA Library Preparation: Life Technologies, Inc.				
	CDNA Library Arrayed by: Greg Lennon, Ph.D.				
	DNA Sequencing by: Washington University Genome Sequencing Center				
	Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:				
	www-bio.llnl.gov/bbrp/image/image.html				
	Insert Length: 1807 Std Error: 0.00				
	Seq primer: -40UP from Gibco				
	High quality sequence stop: 424.				
FEATURES	Location/Qualifiers				
source	1..625				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				

```
/clone="IMAGE:2491091"
/clone_lib="NCI_CGAP_Utl1"
/tissue_type="well-differentiated endometrial
adenocarcinoma, 7 pooled tumors"
/lab_host="DH10B"
/notes="organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Life Technologies catalog #:
11538-014"

BASE COUNT      175 a  155 c  134 g  158 t      3 others
ORIGIN

Query Match.          29.4%; Score 622; DB 10; Length 625;
Best Local Similarity 99.5%; Pred. No. 3e-142;
Matches 622; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1333 aaatacaatcatgaggcctgcttgggtctccagcctggctgctccttctgcgtgatggaagc 1392
Db 625 AAATACAATCATGCGGCTGCTCGGTCTCCAGCCTGGCTGCTTCTGCNTGATGGAAGC 566

Qy 1393 cggatgatggcagtggtgctcctcgtggtgaactctcaacagccagtgagcaggtcgctccc 1452
Db 565 CGGATGATGGCAGTGCGTCCCTCGTGGTGAACCTCTCACAGCCAGTGCGCAGGTCGTC 506

Qy 1453 tctctcctgagacacgacgagtgagacttacttctcatgagttgttcacgtgatgcatt 1512
Db 505 TCTCTCCGACACAGCAGCAGGTGAGGACTTACTTTTCAATGATTTGGTCAGCTGATGCT 446

Qy 1513 cagatttgcacagactgatttgcagatttgcagatttgcagaaacaaatgtggaactgacttt 1572
Db 445 CAGATTTGTGCACAGACTGATTTTGCAGACTTATAGCGGAACAATGTGGAACCTGACTTT 386

Qy 1573 gttaggtgcacatgcgaatgctgaaattgggtgtggagcgtcgattccctccgaaga 1632
Db 385 GTAGAGTGCCTATCGCAAAATGCTTCAAAATTTGGGTGTGGGACGTCGATTTCCCTCCGA 326

Qy 1633 ttgtcaaaacattataagatggaagccctattgcacagactgctgttgaacacttgtt 1692
Db 325 TTGTCAAAACATTATAAGATGGAAGCCCTATTGCAGACGATCTGCTTGAAAACTTGT 266

Qy 1693 gctctaggctggttcaaacacaggttctctgaccctgcccagattgttttgagcaaat 1752
Db 265 GCTTCTAGGCTGGTCAACACAGGCTCTTGTGACCTGGCCAGATTGTTTGTGAGCANAGTT 206

Qy 1753 gatcagctctctcaacacacacacacacacacacacacacacacacacacacacacacac 1812
Db 205 GATCAGTCTCTTCATACCAACACATCGCTGGATGCTGCAAGTGAATATGCCAATACTGC 146

Qy 1813 tcagaaattattaggagttgcagctactccagccacaaatagccagctacacctttggacat 1872
Db 145 TCAGAAATATTAGAGTTGACAGTCTACTCCAGGCACAAATATGCCAGCTACCTTTGGACAT 86

Qy 1873 ttggcaggggatacagatggccaattatttgatatctttggagtgaaagtattttccatg 1932
Db 85 TTGGCAGGGGATACGATGGCCAATATTATGGATATCTTTGGAGTGAAGTATTTTCCATG 26

Qy 1933 gatatttttacagctgttttaaaa 1957
Db 25 GATATGTTTTACAGCTGTTTTAAAAA 1

RESULT 2
BI335570 LOCUS BI335570 789 bp mRNA EST 30-JUL-2001
DEFINITION 60298590F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5140583 5',
mRNA sequence.
ACCESSION BI335570
VERSION BI335570.1 GI:15020227
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 789)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM11346 row: d column: 24
High quality sequence start: 33
High quality sequence stop: 789.
FEATURES
source
1..789
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5140583"
/clone_lib="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/notes="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.4 kb. Library prepared by Life
Technologies."
BASE COUNT      226 a  177 c  211 g  175 t
ORIGIN

Query Match          28.9%; Score 611; DB 11; Length 789;
Best Local Similarity 99.0%; Pred. No. 1.5e-139;
Matches 625; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Qy 1 atgatcgccggtgcttcttggctgctggaagcctccgagagtggtggtccaggatt 60
Db 160 ATGATCGCCCGTGGCTTTTGGCTGTGGGAAGCCTCCGACAGTTGGTGTCCAGGATT 219

Qy 61 ttactcagaatgacgttaggaagaagtgtctctctctctcaggaatgtctctcctat 120
Db 220 TTACTCAGAAATGACGTAGGAAGAGAGTGTCTCTCTTCCAGCAATGTCTTCTCTAT 279

Qy 121 actgtgctggcagaataatgttttaagatggatgtttccaccagagcaaatataaacaaga 180
Db 280 ACTGTGCTGGCAGAAATGTTTAAAGATGGGATCTTTTACCAGAGCAAAATTTAAACAAGA 339

Qy 181 actgagagctcattgtgcagaccacacacaggtgtacgatgctgtggaatgctcggtatt 240
Db 340 ACTGAGGAGCTCATTGTGCAGACCACCAACAGGTGTACGATGCTGTGGNAATGCTCGGTATT 399

Qy 241 gaggaagtaacttacgagaactgtctgcagcactggcagatgtagaagtaaaagtata 300
Db 400 GAGGAAGTAACTTACGAGAACTGTCTGCAGCACTGGCAGATGTAGAAGTAAAGTATATATA 459

Qy 301 gtggaaggagcactgtacgtcttcccccagcatgtatctctctgacaaagaagtcagagca 360
Db 460 GTGGAAGGAGCATTGCTAGACTTTTCCCGCAGCATGTATCTCTGACAAAGAGTACGAGCA 519

Qy 361 gcaagtacagaagcagacacaaagacttctctgttttgatattgagatgagcatgagagga 420
Db 520 GCAAGTACAGAGCAGACACAAAGACTTCTCGTTTGTATTTGATATTGAGATGAGCATGAGAGGA 579

Qy 421 gatattttgagagaattgttcatttacaggaacacctgtgactctggggaagataaaacct 480
Db 580 GATATATTGAGAGAAATGTTTCAATTTACAGAAACCTGTGTATCTGGGAGAGATAAAACCT 639

Qy 481 gagccagagcacttggaagactcaattaaaatgggaaagaaatgggctccatctt 540
Db 640 GAGCCAGACGACTACTTGGAAAAGTCAATTTAAATGGGAAAAGAAATGGGCTCCATCTT 699
```

QY 541 cctgaacaagtacagaatgaatcaaatcaatgaagaaagaatgagtgagctatgtatt 600  
 |||||  
 Db 700 CCTGAACAAGTACAGAATGAAATCAATCAATCAGTGAAGAACAGACAGATGAGTGATGATT 759  
 |||||  
 QY 601 gattttaacaaaactcaatcagagatgata 631  
 |||||  
 Db 760 GA-TTTGACAAAAGACTCAATCAGGATGATA 789  
 |||||

RESULT 3  
 BG180236 579 bp mRNA EST 06-FEB-2001  
 LOCUS 602329846f1 NIH\_MGC\_91 Homo sapiens cDNA clone IMAGE:4431130 5',  
 mRNA sequence.  
 ACCESSION BG180236  
 VERSION BG180236.1 GI:12686939  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 579)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)  
 Tissue Procurement: DCTD/DRP  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1AM10185 row: 1 column: 11  
 High quality sequence stop: 579.  
 Location/Qualifiers  
 1..579  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4431130"  
 /clone\_lib="NIH\_MGC\_91"  
 /tissue\_type="adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: prostate; Vector: pCMV-Sport6; Site:1: NotI;  
 Site:2: SalI; Cloned unidirectionally; oligo-dT primed.  
 Average insert size 1.4 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC Library."  
 BASE COUNT 158 a 117 c 152 g 152 t  
 ORIGIN

Query Match 27.0%; Score 570; DB 11; Length 579;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-129;  
 Matches 570; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1044 gggttttgaatatgaggaatacaatcgctggagatctattactacatgactcagac 1103  
 |||||  
 Db 10 GGGTTTGAATATGAGGAAATCAATGCCCTGGGATCTATATTACTACATGACTCAGAC 69  
 |||||

QY 1104 agaggaactcaagtattccatagaccaaagagttctcctcaagggaatactccccattgaggt 1163  
 |||||  
 Db 70 ACAGGAACCTCAAGTATTCATATAGACACAGAGTCTCTCAAGGAATACTTCCCCAATTGAGGT 129  
 |||||

QY 1164 ggtcactgaagcgttgctgaacacacctaccagagtggttggtggaacttcattggaacaaat 1223  
 |||||  
 Db 130 GCTCACTGAAGCCTTGCTGAACACCTACCAGGAGTGTGTGGACATTCATTTGAACAAAT 189  
 |||||

QY 1224 gacagatgctcattgtttggaacaaagagtggttacatttatactctgtaaggataaagctac 1283  
 |||||  
 Db 190 GACAGATGCTCATGTTGTTGAACAAGAGTGTACACTTTATACATTTATACATGTAAGGATAAAGCTAC 249  
 |||||

QY 1284 agagaagattgggacagattctctatttgagacctctatccaagggaagaaatacaatca 1343  
 |||||  
 Db 250 AGSAGAAGTATTGGGACAGTCTCTATTGGACCTCTATCCAAGSAGGAAGAAATACAAATCA 309  
 |||||

QY 1344 tgcggccttcctcggtctccagcctggctgctctctgctgctgaaagccgagatgagc 1403  
 |||||  
 Db 310 TCGGGCCTGCTTCGGTCTCCAGCCTGGCTGCTTCTGCTGATGGAAGCCGATGATGCC 369  
 |||||

QY 1404 agtgctgcctcctggtggaacttctcacagccagtgccaggtgcctctctctctcag 1463  
 |||||  
 Db 370 AGTGCTGCCCTCGTGGGAACCTTCTCACAGCCAGTGGCAGGCTGCTCCTCTCTCTGAG 429  
 |||||

QY 1464 acacagcaggtgaggaacttacttctcaggttggcagtgatgcacatgatttgctc 1523  
 |||||  
 Db 430 ACACGACGAGGTGAGGACTTACTTTCATGAGTTTGGTACGTTGATGCATCAGATTTGTGC 489  
 |||||

QY 1524 acagactgatttgcagcatttagcgggaacaaatgtagaaactgactttgtagagtgcc 1583  
 |||||  
 Db 490 ACAGACTGATTTTGCACGATTTAGCGGAACAAATGGAAGAACTGACTTTGTAGAGTGCC 549  
 |||||

QY 1584 atcgcaaatgcttgaataattgggtgtgga 1613  
 |||||  
 Db 550 ATCGCAAAATGCTTGAATAATGGGTGTGGA 579  
 |||||

RESULT 4  
 BG666754 902 bp mRNA EST 30-APR-2001  
 LOCUS DRABUG01 Rat DRG Library Rattus norvegicus cDNA clone DRABUG01 5',  
 mRNA sequence.  
 ACCESSION BG666754  
 VERSION BG666754.1 GI:13888661  
 KEYWORDS EST.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 REFERENCE 1 (bases 1 to 902)  
 AUTHORS Xiao,H.S., Han,Z.G., Zhang,F.X., Huang,Q.H., Lu,Y.J., Bao,L., Fu,G.,  
 Guo,C., Yan,Q., Jin,S.X., Zhu,Z.D., Xu,X.R., Li,N.G., Chen,Z. and  
 Zhang,X.  
 TITLE Distinct gene expression profiles of rat dorsal root ganglion  
 induced by peripheral nerve axotomy  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Zhang Xu  
 Laboratory of Sensory System  
 Institute of Neuroscience  
 320 Yue Yang Road, Shanghai 200031, P.R.China  
 Tel: 86-21-64748700-121  
 Fax: 86-21-64713446  
 Email: [xu.zhang@ion.ac.cn](mailto:xu.zhang@ion.ac.cn)  
 This clone is also available at Chinese National Human Genome  
 Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,  
 Pudong New Area, P.R.China. Please contact with Zhang Xu  
 ([xu.zhang@ion.ac.cn](mailto:xu.zhang@ion.ac.cn)) or Han Zeguang ([hanzg@chgc.sh.cn](mailto:hanzg@chgc.sh.cn))  
 PCR Primers  
 FORWARD: T3  
 BACKWARD: T7  
 Seq primer: T3  
 POLYA=No.  
 Location/Qualifiers  
 1..902  
 /organism="Rattus norvegicus"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="DRABUG01"  
 /clone\_lib="Rat DRG Library"  
 /sex="male"  
 /tissue\_type="dorsal root ganglion"  
 /dev\_stage="adult"  
 BASE COUNT 214 a 238 c 239 g 211 t  
 ORIGIN



QY	1570	tttttagaggtgccatcgcgaatgcttgaaattgggtgtgggacgtcgcattccctccga	1629
Db	122	TGTGTAGAGGTGCTTTACAAATGCTTCAAAACCTGGGTGTGGGACATTGACTCCTTGGCGA	181
QY	1630	agattgtcaaacattataaagatggaagccctattgcagacgactcgtttgaaacaatt	1689
Db	182	GAAGTGTCAAAACATCTAGAGACGGACACCCCTATCACAGCAGAGCTGCTGGGACAAAGCTT	241
QY	1690	gttgccttcaggtggtgcaacacaggtctcttgacccctggccagattgtttgagcaaa	1749
Db	242	GTGGCTTCACAGACTGGTCAACACAGAGTCTTTTGACGCTTCGCCAAATGTTTGTGACGAGA	301
QY	1750	gttgatcagctcttcataccaacacacatcgctgtagctgccaagtgaatgccaataac	1809
Db	302	GTTGACACAGCTCCTTCACTACTAATCGTGGGTGGATGCCGAGCGGAATACGCTAAATAC	361
QY	1810	tgcctcagaataattaggagttgcagctactccagcgcaaatatgccagctacctttgga	1869
Db	362	TGCACAGAATTTATGGGAGTTGCAGCTACGCCAGCGCAAAATATGCCAGCTACCTCTGGA	421
QY	1870	catttggcaggggatacgaatggccaaattattatggtatctcttggagtgaggtatttcc	1929
Db	422	CATTTCGAGGAGGGTATCAGCGCCAGTATTATGATATCTGTGGAGTGAAGTCTATTCC	481
QY	1930	atggtatgttttacagctgttttaaaaaaagaggataatgaatccagaggttggaatg	1989
Db	482	ATGGACATGTGCACAGCTGGTTGACACAGAATGGATTATGAATCCAGAGGTTGGAATG	541
QY	1990	aaatacagaacctaatcctgaacctggggatctctgaacgcatgggacatgccac	2049
Db	542	ATATACAGAAACCTAATCTGAAGCTGGGTGTCCTGCACCTCATGGACATGCTCCAG	601
QY	2050	aatttcctgaaactgtag	2067
Db	602	AATTTCTTGCACAGTGAG	619
RESULT	7		
LOCUS	BG3900085	745 bp	EST
DEFINITION	602415924F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4524509 5',		12-MAR-2001
ACCESSION	BG3900085	745 bp	EST
VERSION	BG3900085.1	GI:13283631	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1. (bases 1 to 745)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM10428 row: 0 column: 06 High quality sequence stop: 698. Location/Qualifiers 1. 745 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4524509" /clone_lib="NIH_MGC_92" /tissue_type="embryonal /lab_host="DH10B (phage-resistant)"		
FEATURES			
source			

/note="Organ: testis; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."

BASE COUNT

203 a

166 c

210 g

166 t

Note: this is a NIH\_MGC Library.

ORIGIN

Query Match

22.3%

Score 472.6;

DB 11;

Length 745;

Best Local Similarity

96.3%

Pred. No. 1.6e-105;

Matches 548;

Conservative

0;

Mismatches 14;

Indels 7;

Gaps 6;

Qy

1

atgacgcccgtgctcttggctgtgcgaagcctccgacagtggtggtccaggatt

60

Db

153

ATGATCGCCCGTCCCTTTGGG-TGTGGAAGCCTCCGACAGTTGTTGGTTCCAGGATT

211

Qy

61

ttactcgaatgcagcttaggaagaagtgatgtctctcttcaggcaatgtctcttat

120

Db

212

TTACTCAGAATGACGTAGGAAGAAGTGTCTCTCTCTTCAGGCAATGCTTCCTAT

271

Qy

121

actgtgctgcagaaatgttttaagatgggtatctttcaccagagcaaatataaacaaga

180

Db

272

ACTGTGGCTGCAGAAATGTTTAAAGATGGATCTTTCACCAGACCAATTTAAACAAGA

331

Qy

181

actgagagctcattgtgcagacacaaacaggtgtacgatgctgttggaaatgctcggtatt

240

Db

332

ACTGAGGAGCTCATTGTGCAGACCAACACAGGTGTACGATGCTGTGGAAATGCTCGGTATT

391

Qy

241

gaggaagtaacttcagagaactgtctgcaggaactgtgcagatgtagaagtaaatatata

300

Db

392

GAGGAAGTAACTTACGAGAACTGTCTGCAGCACTGGCAGATGTAGAGTAAAGTATATA

451

Qy

301

gtgaaagaccatgtacgtctccacagcatgtatctctgcacaaagaagtacagaca

360

Db

452

GTGGAAGGACCATGTAGACTTTCCCGAGCATGTATCTCTGCAAGAAGATGACGAGCA

511

Qy

361

gcaagtacagaagcagacacaaagacttctcgttttggattattgagatgagcatgagaga

420

Db

512

GCAAGTACAGAAGCAGACAAAAGACTTCTCGTTTGTATATTGATGAGCATGAGAGGA

571

Qy

421

gata-tattgagagaattgttcattt--acagaaacctgtatctgtggaagataaaaa

477

Db

572

GATATTATTTTCAGAGAAATGTTTCAATTTTACAGGAAACCTGTGATCTGGGAAGATAAAA

631

Qy

478

cctgagccagacatacttgaaaagtcaattaaaaatggggaagaaatgggctccat

537

Db

632

ACTGAGG-CACACGATAC-TGGGAAGTCAATTACATGCGGCAAGAAATGGCGTCCAT

689

Qy

538

cttcctgacaaagtacagaatgaatcaa

566

Db

690

CTT-CTGAACATGTCACGAATGAAATCAA

717

RESULT

8

BE311015

LOCUS

BE311015

559 bp

mRNA

DEFINITION

601088224F1 NCI\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:3483124 5', mRNA sequence.

ACCESSION

BE311015

VERSION

BE311015.1

GI:9171599

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS

1 (bases 1 to 559)

TITLE

NIH-MGC http://mgc.nci.nih.gov/.

JOURNAL

National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys

CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM8513 row: p column: 05  
High quality sequence stop: 557.  
Location/Qualifiers  
1. 559  
/organism="Mus musculus"  
/strain="C57/B6"  
/db\_xref="taxon:10090"  
/clone\_image="3483124"  
/clone\_lib="NCI\_CGAP\_Mam5"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="7 months"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Robin Humphreys, NIH"

BASE COUNT	139 a	136 c	146 g	138 t
ORIGIN				
Query Match	21.5%	Score 455.6;	DB 10;	Length 559;
Best Local Similarity	88.5%	Pred. No. 2.2e-101;		
Matches 494;	Conservative	0;	Mismatches 64;	Indels 0;
Gaps	0;			
Qy	1167	cactgaaggctgtgtgaacacactaccaggaagttgtgggactttcattgacaataaac	1226	
Db	1	CAC7GAAGGCTCTAGACATCTACAGGAGCTGTGGGACTTTCATTTGAACAAGTGC	60	
Qy	1227	agatgctcatgttgggaacagaggtttacactttatactgtgaaggataaagctacagg	1286	
Db	61	TGATGCCCATGTTTGGATAAGAGTGTTCACCTTACACCGTGAAGGATAAAGCTACTGG	120	
Qy	1287	agaagtattggagcagttctatttggacctctatccaaagggaagaaatacaaatcagc	1346	
Db	121	AGAAAGTCTGGGCGAGTTCTACCTGGACCTCTATCCAAGGAAGGAAATAACAACATGC	180	
Qy	1347	ggcctgtcctggtctccagcctgtgctcctctcctgtatgaagccgagatgagcagt	1406	
Db	181	GGCTGCTTTTGGTCTCCAGCCAGGCTGCCCTTCTCCCTGATGGCAGTCGGATGATCTGT	240	
Qy	1407	ggctgacctgtgtgaacttctcacagccagtgccgagtgccctctcctctcctgagaca	1466	
Db	241	GGTGCCCTGGTGGTCAACTTCTCTACCCCATAGCAGCGCGGCCCTCTCTCCTGAGACA	300	
Qy	1467	gcagaggtgaggacttactttcatagtttgggtcacgtgatgcacagatttggacaca	1526	
Db	301	CGAAGTGGCAACTTACTTCCATGAGTTGGTCCATGTCATGATCAGATCTGTGCACA	360	
Qy	1527	gactgattttgcagcatttagcgaaacaaatgtggaactgactttgtagaggtccatc	1586	
Db	361	GACTGACCTTTCACGATTCTAGTGGAAACAAATGTGGAACACTGACTTTGTAGAGTCCATC	420	
Qy	1587	gcaaatgttgaataatgggtgtggagcgtcgattccctccgaaagattgtcaaaacatta	1646	
Db	421	ACAAATGCTTGAACACTGGGTGGGACATTTGACTCCCTGCGAAACACTGTCAAAACATTA	480	
Qy	1647	taaatggaagccctatttgacagcagatctgcttgaataaaccttgttcttctagctgggt	1706	
Db	481	TAGAGACGGACACCTTATCACAGAGGAGCTGCTGGAGAAGCTTGTGGCTTCGAGACTGT	540	
Qy	1707	caacacaggtctcttgac	1724	
Db	541	CAACACAGGTCTCTTGAC	558	
RESULT	9			

AW786845  
LOCUS 120431 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence. EST 09-JUL-2000  
DEFINITION AW786845  
ACCESSION AW786845  
VERSION AW786845.1 GI:7843621  
KEYWORDS EST.  
SOURCE pig.  
ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
REFERENCE 1 (bases 1 to 509)  
Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,  
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.  
and Keele,J.W.  
TITLE Design and use of two pooled tissue normalized cDNA libraries for  
EST discovery in swine  
JOURNAL Unpublished (2000)  
COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@mail.marc.usda.gov  
Single pass sequencing. Bases called and alt\_trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -minmatch 12 options.  
PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCCCGAGTCAGGCG  
Plate: 39 row: N column: 18  
Seq primer: ATTAGGTGACACTATAG.  
FEATURES  
Location/Qualifiers  
1..509  
/organism="Sus scrofa"  
/db\_xref="taxon:9823"  
/clone\_lib="MARC 1P1G"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
Library made from pooled tissue from day 11, 13, 15, 20,  
and 30 embryos."  
BASE COUNT 155 a 99 c 115 g 140 t  
ORIGIN  
Query Match 20.4%; Score 431.4; DB 10; Length 509;  
Best Local Similarity 91.7%; Pred. No. 28-95;  
Matches 467; Conservative 0; Mismatches 41; Indels 1; Gaps 1;  
QY 1515 gattgtgcacagactga-ttttgacgatttagcggaacaaatgtgaaactgactttg 1573  
DB 1 GATTGTGGCGACAGCTCTCTGCAGATTAAATGACACACATGTGGAAACAGACTTTG 60  
QY 1574 tagagtgccatgcgaatgcttgaataattgggtgtgggacgtcgattccctccgaagat 1633  
DB 61 TAGAGGTGCGCATCACAAATGCTTGAACACTGGGTGTGGGACACTGATTCCTCCGAAGGC 120  
QY 1634 tgcataaacattataaagatggaagccctattgcagacgatctgcttgaaaaactgttg 1693  
DB 121 TGTCAAAACATTATAAAGATGGAAGCCCTATTACAGATGATCTGCTGAAAAACTTGTG 180  
QY 1694 ctctaggtgggtcaacacaggtcttctgacctgcgcagattgttttgagcaaatgtg 1753  
DB 181 CTCTAGACTGGTCAACACAGAGTCTCTGACCTTAGCCAGATTGTTTGTAGCAAAAGTTG 240  
QY 1754 atcagctctctacacacacacacacatcgctggatgctgcaagtgaatgcaaatactgct 1813  
DB 241 ATCACCTCTCCACACCAACACTTCGCTGATGCTGCAAGTGAATATGCCAATACTGCA 300  
QY 1814 cagaaatattagaggtgcagctactccaggcacaataatgccagctacacttttggacatt 1873  
DB 301 CAGAAATTTTAGGTGTGCAACTACTCTCAAGCACAAATATGCCAGCTACTTTTGGCACT 360

QY 1874 tggcagggggatcacgattgcccattatttgatatctttggagtgaagtattttccatgg 1933  
DB 361 AGGCAGGGGATGATGATGCCATATATATGATATGTTGGAGTGAAGTCTTTTCCATGG 420  
QY 1934 atatgttttacagctgtgttttaaaagagggaataatccagcaggttggaataaaat 1993  
DB 421 ACATGTGATATACTCTTTTAAAAAAGATGGATAATGAATCCTGAGTTGGATGAAT 480  
QY 1994 acagaaacctaatctgaaacctggggga 2022  
DB 481 ACTGAAACCTAATCTCTGAAACCTGGGGGA 509  
RESULT 10  
LOCUS AW732454  
DEFINITION AW732454 570 bp mRNA EST 21-APR-2000  
similar to SW:NEUL\_RABIT P42675 NEUROLYSIN PRECURSOR ;, mRNA  
sequence.  
ACCESSION AW732454  
VERSION AW732454.1 GI:7632786  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 570)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue Procurement: DCTD/DRP  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
image.llnl.gov/image/html/iresources.shtml  
Seq primer: -40RP from Gibco  
High quality sequence stop: 393.  
FEATURES  
Location/Qualifiers  
1..570  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:2958628"  
/tissue\_type="renal cell adenocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: kidney; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 130 a 138 c 161 g 140 t  
ORIGIN  
Query Match 19.8%; Score 419.8; DB 10; Length 570;  
Best Local Similarity 94.0%; Pred. No. 1.4e-92;  
Matches 436; Conservative 0; Mismatches 28; Indels 0; Gaps 0;  
QY 1 atgatgcccggtgccttttggctgtgcgaagcctccgacagttggtgtccaggatt 60  
DB 106 ATGATGCCCGGTGCCTTTTGGCTGTGCGAAGCCTCGCAGAGTTGGTGTCTCAGGATT 165  
QY 61 ttactcagaatgacgttaggaagaagatgctctctcttcaggcaatgtcttctcat 120  
DB 166 TTACTCAGAATGACGTTAGGAGAGAAGTATGATGTCCTCTTCAGGCAATGTCTCTCAT 225

```
Qy 121 actgtggctgcagaaatgttttaagtggatctttccaccagagcaaatataaacaaga 180
|||||
Db 226 ACTGTGGCTGGCAGAAAATGTTTAAAGATGGATCTTTCACCAGAGCAAAATTAACCAAGA 285
|||||
Qy 181 actgaggagctcattgtgcagaaacacaggtgtacgatgtgttggaatctcggtatt 240
|||||
Db 286 ACTGAGGAGCTCATGTGTGCACACCAACAGGTGTACGATGCTGTGGAAATGCTCGGTATT 345
|||||
Qy 241 gaggaagttaacttacgagaactgtctgcagagcactggcagatgtagaagtaaaagtata 300
|||||
Db 346 GAGGAAGTAACTTACGAGAAGTGTCTGCAGACACTGGCAGATGTGAAGTTAAGTATTTA 405
|||||
Qy 301 gtgaaaggaccatctagactttcccccagcatgtatctctgcacaaagaagtcagagca 360
|||||
Db 406 GTGTGAAGGACCATGCTAGACTTTCCCGCAGCATGTTTCTCTTGCACAGAGGTTCGAGCC 465
|||||
Qy 361 gcaagtacagaagcagacacaaagactttctctgttttgatattgatagatgagagga 420
|||||
Db 466 GCANGGTCAGAAGCAGACAAAAGACTTCTCTGGTTTGTATATTGGGATGGGCTTGGAGGG 525
|||||
Qy 421 gatatttgagaaattgttcattacaggaaacccctgtgatct 464
|||||
Db 526 GATATTTTGAGAGAAATGGTTCATTTTCAGGAACCCCTGGATCT 569
|||||

RESULT 11
AI386253 511 bp mRNA EST 27-JAN-1999
LOCUS mu42g01.y1 Stratagene mouse melanoma (#937312) Mus musculus cDNA
DEFINITION clone IMAGE:524208 5' similar to SW:NEUL_RAT P42676 NEUROLISIN
PRECUSOR ; mRNA sequence.
ACCESSION AI386253
VERSION AI386253.1 GI:4199716
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 511)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterson,R. and Wilson,R.
TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:318056
This read is a RESSEQUENCE of a previously sequenced mouse clone
correct orientation)
This read has been verified (found to hit its original self in the
putative full length read
vector to vector length is 741
Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 351.
FEATURES
source Location/Qualifiers
1..511
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:524208"
/clone.lib="Stratagene mouse melanoma (#937312)"
/tissue_type="melanoma"
/dev_stage="M2 cells"
/lab_host="SOLR (kanamycin resistant)"
/note="Orgn: skin; Vector: pBluescript SK-; Site_1: EcoRI
```

```
; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
dr. From M2 cells, a highly metastatic derivative of the
K-1735 (mouse) melanoma. Average insert size: 1.0 kb;
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCACGAG
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'

BASE COUNT 140 a 115 c 134 g 122 t
ORIGIN

Query Match 19.7%; Score 416.8; DB 10; Length 511;
Best Local Similarity 88.4%; Pred. No. 7.6e-92;
Matches 451; Conservative 1; Mismatches 58; Indels 0; Gaps 0;

Qy 1603 tgggtgtgggacgtcgatccctccgaagattgtcaaaacattataaagatggaagccct 1662
|||||
Db 2 TGGGTGTGGGACATTTACTCCCTCGAAACACTGTCAAAACATTATAGAGAGGGACACCCT 61
|||||
Qy 1663 attcagacgacgtcgttgaataactgttctctcaggtcgtggtcaacacaggtctcttg 1722
|||||
Db 62 ATCACACGACGAGCTTGGGAGAGCTTGTGGCTTCGAGACTGGTCAACACAGGCTTCTTG 121
|||||
Qy 1723 accctgcgcagattgttttgagcaaatgtatcagtcctcttcataccacacacatcgctg 1782
|||||
Db 122 ACGCTTCGCCAAATTTGTTGAGCAAAAGTTGACCAGTCCCTCCATACCAATGCGTCGCTG 181
|||||
Qy 1783 gatctgcaagtgaatatgccaatactcgtctcagaataattagagagttgcagctactcca 1842
|||||
Db 182 GATCGCGGAGGAGTAACGCTAAATACTGCACAGAAATTTTGGGAGTTGCGAGCTACGCCA 241
|||||
Qy 1843 ggcacaaatagccagctacaccttggacatttggcaggggagatcgatggccaattattat 1902
|||||
Db 242 GGCACAAATATGCCAGCTACCTTTTGGACATTTGGCAGGAGGTTATGACGGCCAGTATTAT 301
|||||
Qy 1903 ggatatcttggagtgaagtatttcccatgcatgtttttacagctgttttaaaaaaadaa 1962
|||||
Db 302 GGATATCTTTGGAGTGAAGTCTTTTCCATGGACATGTTTCACAGCTGTTTCAGAAAAGAA 361
|||||
Qy 1963 gggataatgaatccagaggttggaaatgaaatcacagaaacctaatactctgaaacctggggga 2022
|||||
Db 362 GGGATTATGAATCCAGAGGTTTGGAAATGAATACAGAAAACCTAAATCCTGAAGCCTGGGGGG 421
|||||
Qy 2023 tctctggacggcattgacatctcacaatttttgaacctgagcccaacacaaagcgcg 2082
|||||
Db 422 TCCCTGGACGCATGGACATGCTCAGAAATTTCTTGGCAACGTGAGCCAAACACAGAAAGCA 481
|||||
Qy 2083 ttctaatgagtagagcgcctgcagtctycg 2112
|||||
Db 482 TTCCTGATGATCGAGGCGCTGAATGCTTCG 511
|||||

RESULT 12
AW003054/c 757 bp mRNA EST 08-MAR-2000
LOCUS wg61h05.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2475801 3'
DEFINITION similar to SW:NEUL_PIG Q02038 NEUROLISIN PRECURSOR ;, mRNA
sequence.
ACCESSION AW003054
VERSION AW003054.1 GI:5849892
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 757)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
```





```

Qy 1902 tggatatcttggagtgagtgattttccatcgatattgttttacagctgtgttttaaaaaaaga 1961
Db 440 TGGATATCTTTGGAGTGAAGTGTCTTCCATGCACATGTTCCACAGCTGTTTAAAAAAGA 381
Qy 1962 agggataatgaatccagaggttggaatgaataacagaaacacctaactcctgaacctggggg 2021
Db 380 GGGGATCATGAATCCAGAGGTGGAATGAATGAATACAGAAACCTTAATCCCTGAAGCCTGGGGG 321
Qy 2022 atctctgacggcatggacatgctcccaaatctctgaaacgtgagccaaacccaaaaagc 2081
Db 320 GTCCCTGGACGGCATGGACATGCTCCAGAAATTTCTTGCACGTGAGCCAAACCAAAAGC 261
Qy 2082 gttcctaagttagtagcctgcacgtctcgtcgtgta 2115
Db 260 ATTCTAATGAGTCGAGCCCTGAATGGTTCGTAA 227

RESULT 14
BF719641
LOCUS BF719641 675 bp mRNA EST 03-JAN-2001
DEFINITION mab42f07.y1 Soares_NMEBA_branchial_arch Mus musculus cDNA clone
IMAGE:3973020 5' similar to SW:NEUL_RAT P42676 NEUROLYSIN PRECURSOR
;A mRNA sequence.
ACCESSION BF719641 GI:12020643
VERSION BF719641
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 675)
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml

MGI:1473052
Seq primer: -4ORP from Gibco
High quality sequence stop: 457.
Location/Qualifiers
1 . 675
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3973020"
/clone_lib="Soares_NMEBA_branchial_arch"
/tissue_type="branchial arches"
/lab_stage="embryo, 10.5 dpc"
/lab_host="DH10B (phage resistant)"
/notes="vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTACCAACTGTAAGTGGAGCGCGCATGCTATTTTGTGTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT 186 a 161 c 187 g 137 t 4 others
ORIGIN

Query Match 17.68; Score 373.2; DB 11; Length 675;
Best Local Similarity 77.3%; Pred. No. 4.1e-81;
Matches 450; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

```

```

Qy 1 atgatcccegggtcccttttgctgcaagacctccagagtggtgggtccaggatt 60
||||| ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94 ATGATCACCTCTGCTTTCGGCTCTGCGAGCCCTCCACAGAGCTGGTGGTCCAGGATT 153
Qy 61 ttactcgaatgaacgttttaggaagaagtgatgtctctctctcaggaacaattcttcctat 120
||||| ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 154 CGGTCAGAATGACGTTTGGGAGGAGGCGGCATCTCCCTCTTCAGGCCATGCTTCCTTAC 213
Qy 121 actgtgctgcacaaatgttttaagtgagctcttcaccagagcaaatataaacaaga 180
||||| ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 214 ACCCGCGCAGCAGGAATGTTCTCAGATGGGACCTTTCTCCAGAGCAGATCAGACGAGG 273
Qy 181 actgaggagctcattgtgcagaccacacaggtgtacgatctgtgtggaaatcctcggtatt 240
||||| ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 274 ACAGAGGAGCTCATAGCACACACCAAGCAGGTGTATGATACCGTGGGACCAATCAACCTG 333
Qy 241 gaggaagtaacttacgagaactgtctcaggaactggcagatgtagaagtaaaagtata 300
||||| ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 334 GAGGATGTAACCTACGAGAACTGTCTGAGGTGCTGGCTGACATAGAAGTGAAGTACATA 393
Qy 301 gtgaaaggacactcgtacactttcccacagcatatcctctgacaaagaagtagcagca 360
||||| ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 394 GTGGAAGGACCATGCTGGACTTCCCTCAGCACAGTGTCTGTGACAGAGAAGTGGCGGCT 453
Qy 361 gcaagtacagaagcagacacaaagacttctctgttttattgatatgtgagatgagcatgagga 420
||||| ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 454 GCAAGCAGACAGCGGACAAAAGGCTGTCTGTTGTATATTGAGATGAGCATGAGAGAA 513
Qy 421 gatatttgagaagaattgttcatttacaggaacacctgtgatctcgtgggaagaataaacct 480
||||| ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 514 GATGATTTTCAGAAAATAGGTGTCATGTACAAGAACGTCGCAATTTGGAGAAGATAAAGCT 573
Qy 481 gagccagacatacttgaaaagtcaattaaaatggggaagaaatgggctccatctt 540
||||| ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 574 GAGCCAGGCGCATCTTTGGAGAAATCAATTAANATGTTAAAGGATGGAGCTCATTTT 633
Qy 541 cctgaacaagtacagaatgaaatcaatcaatcaatgaagaaga 582
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 634 CAAACACNTTATAANATGAATTNAATTAATGAATAGAGAAA 675

```

## RESULT 15

```

AA152501
LOCUS AA152501 417 bp mRNA EST 01-AUG-1997
DEFINITION z007a04.r1 Stratagene endothelial cell 937223 Homo sapiens cDNA
Clone IMAGE:568958 5' similar to SW:NEUL_RABIT P42675 NEUROLYSIN
PRECURSOR ;, mRNA sequence.
ACCESSION AA152501
VERSION AA152501.1 GI:1718695
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 417)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E.,
Woodward,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
TITLE
JOURNAL
MEDLINE
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

```











Db 150 cccaggaaacacgtctgcgaacacctcgccatctggcaggtgggtaccacgcccagt 209  
QY 1898 attatgatatctttggagtgaaagtattttccatggtatgttttacagctgttttaaaa 1957  
Db 210 actacgggtacctgtggagcgaggtgtattccatggacatgtttccacacgcgtttcaagc 269  
QY 1958 aagaaggataatgccacccagaggttgaaatgaaatcacagaaacaaatctgaaacctg 2017  
Db 270 agaggggctgctgaaacagcaaggttggcatggattacagaaagctgcatactgagacccg 329  
QY 2018 ggggatctctggcggcatggacatgctccacaatttcttgaacgtgagccaaacaaa 2077  
Db 330 gcggttccggagtcgacgcgcctatgctggcgcttctctggcgctgaccccaagcagg 389  
QY 2078 aagcgttccctaatgagtagaggcctgcagt 2107  
Db 390 acgcttctctctgagcaaggggctgcagg 419

## RESULT 7

US-09-933-524A-7235

; Sequence 7235, Application US/09933524A

; GENERAL INFORMATION:

; APPLICANT: Drmanac, Radoje T.

; APPLICANT: Labat, Ivan

; APPLICANT: Stache-Crain, Birgit

; APPLICANT: Dickson, Mark

; APPLICANT: Jones, Lee W.

; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained

; TITLE OF INVENTION: From Various Libraries

; FILE REFERENCE: 774

; CURRENT APPLICATION NUMBER: US/09/933,524A

; PRIOR FILING DATE: 2001-08-20

; PRIOR APPLICATION NUMBER: 09/528,409

; NUMBER OF SEQ ID NOS: 116231

; SOFTWARE: Hy-patent.pl Version 3.1

; SEQ ID NO 7235

; LENGTH: 468

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc.feature

; LOCATION: (1)...(468)

; OTHER INFORMATION: n = A,T,C or G

US-09-933-524A-7235

Query Match 5.8%; Score 121.8; DB 6; Length 468;  
Best Local Similarity 64.2%; Pred. No. 1.7e-22;  
Matches 183; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 128 ctggcagaaatgttttaagatggatctttccacagagcaaaattaaacagaagactgagg 187  
Db 184 ctgtgttaacagacctgcgtgggacacctgagtcgccacagatagagagcgccacagg 243  
QY 188 agctcatgtgcagaccacaaacaggtgtgacgagctgttggaatgctgctgtattgaggag 247  
Db 244 agctcatgcagacacaaacagcgctgtatgaccaggttggcaccagagtttgaggagcg 303  
QY 248 taacttacgagaactgtctgcagccactggcagatgtgagaagtaagtatatagtggaag 307  
Db 304 tgtctacagagacgcgtcaagcgctggccgagtggtgaggctacacacagttcaga 363  
QY 308 ggaacatgctagactttcccccagacgtatctctgtgacaagaagtagacagcagaagta 367  
Db 364 ggaatctctgtacttccccagcagcatgttccccctccaaagacatccggacagccagca 423  
QY 368 cagaagcagacaagaagactttctgtttgtatgtatgagatgagca 412  
Db 424 cagaggccgacaagaagctctctgaggttcgagctggagatgagca 468

## RESULT 8

US-09-388-906A-20678

; Sequence 20678, Application US/09388906A

; GENERAL INFORMATION:

; APPLICANT: Havukkala, Ilkka

; APPLICANT: Shenk, Michael Andrew

; TITLE OF INVENTION: Polynucleotides Isolated From Plants and

; TITLE OF INVENTION: Methods For Their Use

; FILE REFERENCE: 11000.1013U

; CURRENT APPLICATION NUMBER: US/09/388,906A

; CURRENT FILING DATE: 1999-09-01

; NUMBER OF SEQ ID NOS: 24843

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 20678

; LENGTH: 511

; TYPE: DNA

; ORGANISM: Eucalyptus grandis

US-09-388-906A-20678

Query Match 4.7%; Score 98.6; DB 6; Length 511;  
Best Local Similarity 52.0%; Pred. No. 3e-16;  
Matches 244; Conservative 0; Mismatches 224; Indels 1; Gaps 1;

QY 1077 ggatctatattactacatgactcagacagagggaaactcaagtattccatagaccagagtt 1136  
Db 35 ggatctatgtattattataagacatttcaagagcagcactttgatctatgattttggagc 94  
QY 1137 cctcaagaataacttcccaattgaggtggtcactgaaggttctgtaaacctaccagga 1196  
Db 95 cattagcaatattttccggtcaaatgattctctatcagggcatctttaaattttcccaaga 154  
QY 1197 gtgttgggactttcatttgaacaaa-tgacagatgctcatgttttggacaagagtgtta 1255  
Db 155 cctatttgggttgagatttgaggaaattgccgacctgaggtctggtcattgtgatgtc 214  
QY 1256 cactttatctgtgaaggataaagctacagaggaagatttgggacagttcttatttggacc 1315  
Db 215 gattatttccagttattgacttaagtccggtgactctcctctggggcatttctaccttgata 274  
QY 1316 tctatcaaggaggaagaaatacaatcatcgcctgctctcagctgctcagctggcgcc 1375  
Db 275 tgtatagaagaagaaagacggtcacatcgcgtggttgcctcttcacagacagatcat 334  
QY 1376 ttctgcctgatggaaagccggatgatggcagtggtgctgcctcgtggtgaaacttctcaagc 1435  
Db 335 tatcttttaacaatgttggccagataccagtggtatgtgtaataactcaattcccaagg 394  
QY 1436 cagtggcaggtcgtccctctctctctctctctctctctctctctctctctctctctctct 1495  
Db 395 atatcagtggtcactcctctctctctctctctctctctctctctctctctctctctctct 454  
QY 1496 ttgtcagctgatacctcagatttggcacagactgattttggcagatt 1544  
Db 455 ttggacatgtgtcagcagatatactctgcaacagggcatctatttggcagatt 503

## RESULT 9

US-09-933-524A-3065

; Sequence 3065, Application US/09933524A

; GENERAL INFORMATION:

; APPLICANT: Drmanac, Radoje T.

; APPLICANT: Labat, Ivan

; APPLICANT: Stache-Crain, Birgit

; APPLICANT: Dickson, Mark

; APPLICANT: Jones, Lee W.

; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained

; TITLE OF INVENTION: From Various Libraries

; FILE REFERENCE: 774

; CURRENT APPLICATION NUMBER: US/09/933,524A

; CURRENT FILING DATE: 2001-08-20

; PRIOR APPLICATION NUMBER: 09/528,409



0:



; OTHER INFORMATION: unsure at all n locations  
; OTHER INFORMATION: Clone ID: LIB3587-284-Q1-K6-C10  
US-09-865-439A-67646

```
Query Match          3.7%; Score 78.2; DB 5; Length 637;
Best Local Similarity 52.2%; Pred. No. 1e-10;
Matches 198; Conservative 0; Mismatches 178; Indels 3; Gaps 1;

QY 1396 atgatggcagtgccctcgtgtggaacttctcacagccagtggcaggtcgccctct 1455
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 184 aggcgcctgttgcccatatagtgtgcaatcagactccaccagttggcgagaagccagt 243
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1456 ctccctgagacacgacgaggtgaggacttactttcatgagtttggtcacgtgatgcacag 1515
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 244 ctcatgacctccgtgaggttgaaactgtgtcccatgaatttgggcacgccccttcagcac 303
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1516 atttgtcacagactgattttgcacgatttagcggaacaaat---gtgaaactgacttt 1572
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 304 atgcttactaacaagatgaaggctttgtgtgctgcattcgtggagtgaatgggatgct 363
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1573 gttagaggtgccatcgcaaatgttgaaattgggtgtgggacgtcgattccctccgaaga 1632
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 364 gttagaggtaccctccagttcatgaggaactggtgctatcacagaataactcttttgagc 423
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1633 ttgtcaaaacattataagatggaagccctattgcagacgactctgttgaaaaacttgtt 1692
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 424 attgcaaaacattatgaaccggtgaaacccctccagaggaatttatgcgaagcttgta 483
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1693 gcttctaggtggtcaacacaggtcttctgacctgcgcagattgttttgagcaaat 1752
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 484 gctgcaagaatttccgtgctggccacttgagcctgagcctgcgtcagatacagatttgcagtgta 543
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1753 gatcagttcttcatacca 1771
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 544 gatagggagcttcatacaa 562
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: January 14, 2002, 21:26:15  
Job time: 6724 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 18:43:21 ; Search time 18.52 Seconds  
(without alignments)  
855.417 Million cell updates/sec

Title: US-09-833-782-2

Perfect score: 3668

Sequence: 1 MIARCLLAVSRRLRVGGRI.....FLKREPQKAFMLMRLHAP 704

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA.\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pap.\*

2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pap.\*

3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pap.\*

4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pap.\*

5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pap.\*

6: /cgn2\_6/ptodata/2/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2979	81.2	709	5	PCT-US92-00731-13
2	2331	63.5	689	1	US-07-766-351-5
3	2331	63.5	689	1	US-08-059-032-5
4	2331	63.5	689	5	PCT-US91-07290-5
5	122	3.3	3248	1	US-08-353-700-1
6	122	3.3	3248	5	PCT-US95-16216-1
7	120	3.3	1388	2	US-08-685-576-1
8	116	3.2	892	1	US-07-977-434-12
9	116	3.2	892	1	US-08-458-819-12
10	116	3.2	892	5	PCT-US91-07035-12
11	115	3.1	477	4	US-09-128-142-2
12	115	3.1	505	4	US-09-128-142-4
13	115	3.1	872	1	US-08-766-014-2
14	113.5	3.1	643	2	US-08-245-511-47
15	113.5	3.1	643	2	US-08-600-993A-47
16	113	3.1	976	4	US-09-104-324B-4
17	113	3.1	1388	2	US-08-685-576-4
18	112.5	3.1	615	4	US-08-989-299-11
19	112	3.1	1164	4	US-08-923-992A-2
20	112	3.1	2285	4	US-09-308-375-2
21	112	3.1	2496	4	US-09-125-028-2
22	111.5	3.0	712	2	US-08-468-576B-17
23	111.5	3.0	712	2	US-08-468-576B-17
24	111.5	3.0	712	3	US-08-468-577B-17
25	111.5	3.0	893	2	US-08-706-702-3
26	111.5	3.0	893	3	US-08-706-706-3
27	111.5	3.0	1713	3	US-08-600-982-24

28	111.5	3.0	1713	5	PCT-US94-10261A-24	Sequence 24, Appl
29	111	3.0	1354	3	US-08-685-871-2	Sequence 2, Appl
30	111	3.0	1786	4	US-08-973-462-8	Sequence 8, Appl
31	110.5	3.0	422	4	US-08-646-693-2	Sequence 2, Appl
32	110.5	3.0	422	5	PCT-US96-06053-2	Sequence 2, Appl
33	110	3.0	351	1	US-08-402-217A-2	Sequence 2, Appl
34	110	3.0	351	1	US-08-700-178-2	Sequence 2, Appl
35	110	3.0	351	3	US-08-995-654-2	Sequence 2, Appl
36	110	3.0	2958	4	US-08-894-344C-2	Sequence 2, Appl
37	109.5	3.0	476	4	US-08-134-557D-2	Sequence 2, Appl
38	109.5	3.0	788	2	US-08-907-166-6	Sequence 6, Appl
39	109	3.0	475	2	US-08-484-200-2	Sequence 2, Appl
40	109	3.0	475	3	US-08-465-375-2	Sequence 2, Appl
41	109	3.0	1128	4	US-08-923-992A-6	Sequence 6, Appl
42	107.5	2.9	1264	1	US-07-789-915A-6	Sequence 6, Appl
43	107.5	2.9	1264	1	US-08-005-002C-6	Sequence 6, Appl
44	107.5	2.9	1264	1	US-08-487-203A-6	Sequence 6, Appl
45	107	2.9	1698	4	US-09-315-793-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1  
PCT-US92-00731-13  
; Sequence 13, Application PC/TUS9200731  
; GENERAL INFORMATION:  
; APPLICANT: Kawabata, Shunichiro  
; APPLICANT: Davie, Earl W.  
; TITLE OF INVENTION: MICROSOMAL ENDOPEPTIDASE  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed and Berry  
; STREET: 6300 Columbia Center  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: US  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/00731  
; FILING DATE: 19920128  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/646,997  
; FILING DATE: 28-JAN-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 990008.415PC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206)622-4900  
; TELEFAX: (206)682-6031  
; TELEX: 3723836  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; TYPE: AMINO ACIDS  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: N-terminal  
PCT-US92-00731-13

Query Match 81.2%; Score 2979; DB 5; Length 709;  
Best Local Similarity 86.7%; Pred. No. 4.6e-264;  
Matches 568; Conservative 29; Mismatches 56; Indels 2; Gaps 1;

Qy 2 IARCLAVRSRLRRYGGSRILLRMTLGRVMSPLQAMSSYTVAGRNVLRWLDSPEQIKTRT 61  
Db 54 IARCSAARGHLRHVGGSRILLKMTLGRVMSPLQAVSSYTAAGRNVLRWLDSPEQIKTRT 113  
Qy 62 BELIVQKQYDVGMLGIEBVTYENCLOALADVEVYIIVERTMLDPQHVSSDKVRAA 121  
Db 114 BELIAQTKQVYDSVGMLDIKDVTYENCLOALADVEVYIIVERTMLDPQHVSTDRVRAA 173  
Qy 122 STEADKRLSRDIEMSMRGDIFFERIVHLQETCDLKGKPKPEARRYLEKSIRKNGHLPL 181  
Db 174 STEADKRLSRDIEMSMRGDIFFERIVHLQETCDLKGKPKPEARRYLEKSIRKNGHLPL 233  
Qy 182 EQVQNEIKSMKRMSELICIDFNKNLNNEDDTFLVFSKAELGALPDDFIDLSLEKTDYKI 241  
Db 234 XEVQNXIKSMKRMSELICIDFNKNLNNEDDTFLVFSKAELGALPDDFIDLSLEKTDYKI 293  
Qy 242 TLKYPHYFPVNMKCCIPETRRRMEAFNTRCKEENTILQQLLPLRTYKAKLLGYSHAD 301  
Db 294 TLKYPHYFPVNMKCCIPETRRRMEAFNTRCKEENTILQQLLPLRTYKAKLLGYSHAD 353  
Qy 302 FVLEMTAKSTSRVTAFLDLSQKLPGLGEAREFIFLNKKCKDRGFEDYDGINAWDL 361  
Db 354 FVLEMTAKSTSRVTAFLDLSQKLPGLGEAREFIFLNKKCKDRGFEDYDGINAWDL 413  
Qy 362 YYYMTQTEELKYSIDQEFLEKEYFFIEV--VTGELLNTYQELLGLSFQMTDAHVNMKSVT 419  
Db 414 HYYMTQTEELKYSIDQEFLEKEYFFIEV--VTGELLNTYQELLGLSFQMTDAHVNMKSVT 473  
Qy 420 LYTVKDKATGEVLGQFYLDLYPREGKYNHAAACFGLQPCGLLPDGRMMAVAVVNSQP 479  
Db 474 LYTVKDKATGEVLGQFYLDLYPREGKYNHAAACFGLQPCGLLPDGRMMAVAVVNSQP 533  
Qy 480 VAGRPSLLRHDEVITYPHEFGVHMQICAOQDFARFSGTNNVETDFEVPVNSOMLENWYDV 539  
Db 534 VAGRPSLLRHDEVITYPHEFGVHMQICAOQDFARFSGTNNVETDFEVPVNSOMLENWYDV 593  
Qy 540 DSLRRLSKHYKDGSPDIADLLEKLVASRLVNTGLTLRQIVLSKVDSQSLHTNTSLDAASE 599  
Db 594 DSLRRLSKHYKDGSPDIADLLEKLVASRLVNTGLTLRQIVLSKVDSQSLHTNTSLDAASE 653  
Qy 600 YAKYCEITLGAAPGTNNPATFGLAGGYDQGYGYLWSEVFSMDMYSCFKKE 654  
Db 654 YARYCTDILGAAPGTNNPATFGLAGGYDQGYGYLWSEVFSMDMYSCFKKE 708

## RESULT 2

US-07-766-351-5

; Sequence 5, Application US/07766351

; Patent No. 5292852

; GENERAL INFORMATION:

; APPLICANT: Sinha, Sukanto

; APPLICANT: Seubert, Peter A.

; APPLICANT: Dovey, Harry F.

; APPLICANT: McConlogue, Lisa C.

; APPLICANT: Little, Sheila P.

; APPLICANT: Johnstone, Edward M.

; TITLE OF INVENTION: Amyloidin Protease and Uses Thereof

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Athena Neurosciences, Inc.

; STREET: 800F Gateway Blvd.

; CITY: South San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07766,351

; FILING DATE:

; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy, Lisabeth Feix  
; REGISTRATION NUMBER: 31547  
; REFERENCE/DOCKET NUMBER: 17796-002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 877-0900  
; TELEFAX: (415) 877-8370  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 689 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
; US-07-766-351-5

Query Match 63.5%; Score 2331; DB 1; Length 689;

Best Local Similarity 64.9%; Pred. No. 9.5e-205;

Matches 426; Conservative 98; Mismatches 132; Indels 0; Gaps 0;

Qy 46 NVLRWDLSPQIKTRTEBELIVQKQYDVGMLGIEBVTYENCLOALADVEVYIIVERTM 105  
Db 22 NDLRWDLSPQIKTRTEBELIVQKQYDVGMLGIEBVTYENCLOALADVEVYIIVERTM 81  
Qy 106 LDFPQHVSSDKVRAAASAEADKRLSRDIEMSMRGDIFFERIVHLQETCDLKGKPKPEARR 165  
Db 82 LDFPQHVSSDKVRAAASAEADKRLSRDIEMSMRGDIFFERIVHLQETCDLKGKPKPEARR 141  
Qy 166 LEKSIKMGKRNGLHLPEQVQNEIKSMKRMSELICIDFNKNLNNEDDTFLVFSKAELGALPD 225  
Db 142 LERLIKLRNGLHLPRETQENIKRIKKLSLLCICIDFNKNLNNEDDTFLVFSKAELGALPD 201  
Qy 226 DFIDSLKTDGKYLITLKYPHYPMKCCIPETRRRMEAFNTRCKEENTILQQLLPL 285  
Db 202 DFLNLSKMGDKLKVTLKYPHYPMKCCIPETRRRMEAFNTRCKEENTILQQLLPL 261  
Qy 286 LRTVAKLLGYSHADFVLEMTAKSTSRVTAFLDLSQKLPGLGEAREFIFLNKKCKDRGFED 345  
Db 262 LRAQSRLLGLFHTADYVLENNMAKTSQVATFDELAKLPLGEQERAVILELKRAC 321  
Qy 346 KDRGFEDYDGINAWDLLEKLVASRLVNTGLTLRQIVLSKVDSQSLHTNTSLDAASE 405  
Db 322 ERRGLPFDGRIRAWDMRYMNMQVEETRYCVDQNLKKEYFPVQVWTHGLGIYQELLGLAF 381  
Qy 406 EQMTDAHVNMKSVTLYTVKDKATGEVLGQFYLDLYPREGKYNHAAACFGLQPCGLLPDGR 465  
Db 382 HHEGASAWHEDVRLYTARDAASGEVVGKFLDLYPREGKYNHAAACFGLQPCGLLPDGR 441  
Qy 466 MNAVALVYVNSQPVAGRPSLLRHDEVITYPHEFGVHMQICAOQDFARFSGTNNVETDFV 525  
Db 442 QIATAAMVANETKPTADAPSLQHDDEVITYPHEFGVHMQICAOQDFARFSGTNNVETDFV 501  
Qy 526 EVPSOMLENWYDVDSLRLSKHYKDGSPDIADLLEKLVASRLVNTGLTLRQIVLSKVDS 585  
Db 502 EAPSOMLENWYDVDSLRLSKHYKDGSPDIADLLEKLVASRLVNTGLTLRQIVLSKVDS 561  
Qy 586 QSLHTNTSLDAASEYAKYCEITLGAAPGTNNPATFGLAGGYDQGYGYLWSEVFSMD 645  
Db 562 QALHTQTDADPAEEYARLCOEILGVPAFTPGTNNPATFGLAGGYDQGYGYLWSEVFSMD 621  
Qy 646 MFYSCFKKEGIMNPEVGMKRNILKPGGSLDGMMDLHNLKREPKNQAKFLMSRGL 701  
Db 622 MFHTFKEGVLNLSKVGMDYRSCILRPGGSEDAASAMLRREFLGRDPKQDAFLLSKGL 677

## RESULT 3

US-08-059-032-5

; Sequence 5, Application US/08059032

; Patent No. 5424205

; GENERAL INFORMATION:

[illegible]

```

1 CLASSIFICATION: 435
2
3 ATTORNEY/AGENT INFORMATION:
4
5 NAME: Murphy, Lisabeth Feix
6
7 REGISTRATION NUMBER: 31547
8
9 REFERENCE/DOCKET NUMBER: 17796-002
10
11 TELECOMMUNICATION INFORMATION:
12
13 TELEPHONE: (415) 877-0900
14
15 TELEFAX: (415) 877-8370
16
17 INFORMATION FOR SEQ ID NO: 5:
18
19 SEQUENCE CHARACTERISTICS:
20
21 LENGTH: 689 amino acids
22
23 TYPE: AMINO ACID
24
25 TOPOLOGY: linear
26
27 MOLECULE TYPE: protein
28
29 HYPOTHEetical: NO
30
31 ANTI-SENSE: NO
32
33 PCT-US91-07290-5

```

**Qy** 46 NVLRWDLSPSEQIKRTEELIYQTKQVYDAVGMLGIEEVTYENCILQALADVEVKYIVERTM 105  
- - - - - : : : - - - - - : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
**Dd** 22 NDLRWDLSAQQIEETRELISQTKRKYDVQGTGFEDVSVESTUKALADVEVTTYVQRNI 81

Qy 106 LDFQHVSDREKRAASTADKRLSRFDIEMSRGIDIPERIVHQLQETCDLGKIKPEARRY 165  
Db 82 LDFQHVSPSDIRTAETADKLSSEFVEMREDYQRIWVLQEKVKQKDSLREPEARRY 141  
Qy 166 LKSIKMGKRNGLHLPQGVQVNEIKSMKRMBELCIDFNKNLNEDDTFLVFSKAEALGALPD 225  
Db 142 LERLIKLGRRNGLHLPRETOENIRIKKLSLLCICDFNKNLNEDTFLPFTLQELGGLPE 201  
Qy 226 DFIDSEKTDGKYLITLYKYPHYPMVKKCCIPETRRRMEAFNTRCKEENTIILOQLLP 285  
Db 202 DFLNLEKMDGKLUKLUKYPHYPLKAKCHVPETRRRVEAFNCRKEENAILKELVT 261  
Qy 286 LRTVAKLLGYSTHADFVLEMTAKTSRVTAFLLDLSQKLPGLGEAREFILNKKKEC 345  
Db 262 LRAQSRLLGFHTADYVLENNMAKTSQVATFELDELAQKLPGLGEQRAVILELKRAC 321  
Qy 346 KDRGFYDCKINAWDLYYMTQTELYKSIOEFLKEFPEVTEVTEGLNYYQELLGSLF 405  
Db 322 ERRLGPDGRIRAWDMRYMNMQVEETRYCDONLLKEYFPVQVTVHGLLGYQELLGLAF 381  
Qy 406 EQMTDAHVWNKSVTLTYTKDKATGEVLGQFLDLYPREGKYNHAAACFGQLPGCLLPDGSR 465  
Db 382 HHREGASAHEDVRLYRTADAASEVGVKFLDLYPREGKYGHACFGQLPGCLLPDGSR 441  
Qy 466 MVAALVYVNSQPVAGRPSLLRHDEVRTYFHEFGHVMHQICATQDFARFSGTNVETDFV 525  
Db 442 QIATAAMVANFTKPTADAPSLQLHDEVEYFHEFGHVMHQLCSQAEPAMFSGTHVERDFV 501  
Qy 526 EYPSQMLNWWVDLSRLSKHYKDGSPDIADDLLEKLVASRLVNTGLTLTRQIVLSKVD 585  
Db 502 EAPSOMLNWWYEQEPLLRMSRHYRTGSAVPRELLEKLVESQANTGLFLSRQIVLAKVD 561  
Qy 586 OSLHTSLDAASEYAKYCSILGVAAATPGTNMPATFCHLAGGYDGOYGYLWSEVFSMD 645  
Db 562 QALHTQDADPAEYARLCQILGVATPGTNMPATFCHLAGGYDAQYGYLWSEVFSMD 621  
Qy 646 MFYSCFKKEGIMNPEVGMKYNRLNLPKPGSLDGMDFHNLKREPQKAFMLSRGL 701  
Db 622 MFHTRFQEGVLNSKGVMDYRSCILRPGGSEDASAMLRFLGRDPKQDAFLLSKGL 677

RESULT 5

US-08-353-700-1  
; Sequence 1, Application US/08353700  
; Patent No. 5599919  
; GENERAL INFORMATION:  
; APPLICANT: YEN, TIMOTHY J.  
; APPLICANT: RATTNER, JEROME B.  
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A  
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,  
; TITLE OF INVENTION: AND METHODS OF USE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN  
; STREET: 1601 MARKET STREET, SUITE 720  
; CITY: PHILADELPHIA  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103-2307  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/353,700  
; FILING DATE: 09-DEC-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: REED, JANET E.  
; REGISTRATION NUMBER: 36,252  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 563-4100

; TELEFAX: (215) 563-4044  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3248 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORGANISM: HUMAN  
; US-08-353-700-1

Query Match 3.3%; Score 122; DB 1; Length 3248;

Best Local Similarity 19.8%; Pred. No. 0.22; 238; Indels 142; Gaps 25;  
Matches 116; Conservative 91; Mismatches 238; Indels 142; Gaps 25;  
Qy 56 QIKRTTEELIVQTKQYDAVGMGLGIEEVTYENCQALAD--VEVKYIVERTMLDPPQHVS 113  
Db 416 QMKARLTQELQAKNMH-----NVLQALDKLTSVKQOLENNLEEFKQKLC 461  
Qy 114 SDKEVRAASTADKRLSRFDIEMSRGIDIPERIVHQLQETCDLGKIKPEARR--YLEKSIIK 171  
Db 462 RAEQAFQASQIKENELRRSMEEMKENLLKS--HSEQ-----KAREVCHLEAEIK 510  
Qy 172 MGRNGLHLPEQVQNEIKSMKRMBELCIDFNKNLNEDDTFLVFSKAEALGALPDIDSL 231  
Db 511 NIKQ--CLNQSNQFAEEMKAKNTSQETMLRDLOEKINQENSLTLEKLA-----VADL 563  
Qy 232 EKTDDDKYKILTKYPHYPMVKKCCIPETRRRMEAFN-----RCREENTII-- 280  
Db 564 EKQDCSDLLKKREHTEQLND-KLSKTEKESKALLSALELKKKEYBELKEETLFSOW 622  
Qy 281 -----QQLLPRTKVAKLLGYSTHADFVLEMTAKTSRVTAFLLDLSQKLPGLGEARE 335  
Db 623 KSENEKLLTOMESEKENLQSKINHLETCCLKTQOIKS-----HEYNERVTL-EMDRE 673  
Qy 336 FILNLKKCEKDRGFYDCKINAWDLYYMTQTELYKS-----IDQFLKKEYFPTE 387  
Db 674 ---NL-----SVEIRNLNVLDSKSVETQKLAWELOQKAEFSDQKHQKIEHNC 722  
Qy 388 VVTEGLLNTYQEL---LGLSFEQMTDAHVWNKSVTLTYTKDKATGEVLGQFYLDLYPREG 444  
Db 723 LKTSQLTQGVDELHKLQLLSNEIMD-----KDRC-----YQDL----- 756  
Qy 445 KYNHAAACFGQLPGCLLPDGSRMMAVAALVYVNSQPVAGRPSLLRHDEVRTYFHEFGHVMH 504  
Db 757 ---HAEYESLRDLKLSKD-----ASLVTNEDH---QRSLLAFDQOPAMHHSFANIIG 802  
Qy 505 QICAQTDFARESGTVNVTDFVEVPSQMLNWNVDYDLSR-----RLSKHYKDG 553  
Db 803 EOGSNP---SERSECRLEADQSPKNSAILQN---RVDSEFLESOKQMNLSLQCOCELV 857  
Qy 554 PIADDLLEKLVASRLVNTGLL--TLRQIVLSKVDQSLHTNTSLDAAS 598  
Db 858 QIKGEIEENLMAEQMHQSFAETSQRISKLEQEDTSAHQNVVAETLS 904

RESULT 6

PCT-US95-16216-1  
; Sequence 1, Application PC/TUS9516216  
; GENERAL INFORMATION:  
; APPLICANT: Yen, Timothy J.  
; APPLICANT: Rattner, Jerome B.  
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently  
; TITLE OF INVENTION: Expressed Kinetochore Protein, and Methods of Use  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman  
; STREET: 1601 Market Street Suite 720  
; CITY: Philadelphia





Db 721 IYSIEBAKSEAMKEMKKLSEBTLKQKVENLLEAEKRCSTLDCDLKOSQKI-----N 776  
Qy 226 DFIDSLEKTDYKIKPLKYPFVPMKKCCIPETRRMEMAFNT-----RCKEENTII 279  
Db 777 ELLKQKDVNEDVRLNFKIEQ--ETQKRCITQNDLKMOTQVNTLKMSEKQLKQENHIL 834  
Qy 280 LQQLPLRTKAKLLGYSTHAD-----FVLENTAKST---SRVTAFLDLSOKLKL 329  
Db 835 LEMKMSLEKQNAELKQERQADGOMKELQDQLEAEQYFSTLYKTQVRELKBECEKTKLC 894  
Qy 330 GEABREILNKKKECDRGFEYDGGKINAWDLYYMTQTEELKYSIDQEFLEKVEYPIEV 389  
Db 895 KE-----LOOKKELQD---ERDSLAQAQELTUTKADSOARSIAEE---QISDLE-- 940  
Qy 390 TEGLLNTYQ--ELGLSFEQMTDAHVWNKSVTLTYVKD 425  
Db 941 KEKIMKELEIKEMMARHKQELTE-----KDATIASLEE 973  
RESULT 8  
US-07-977-434-12  
; Sequence 12, Application US/07977434  
; Patent No. 5466591  
; GENERAL INFORMATION:  
; APPLICANT: Gelfand, David H.  
; APPLICANT: Abramson, Richard D.  
; TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF  
; TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES  
; NUMBER OF SEQUENCES: 38  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Hoffmann-La Roche Inc.  
; STREET: 340 Kingsland Street  
; CITY: Nutley  
; STATE: New Jersey  
; ZIP: 07110-1199  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: 7  
; SOFTWARE: WordPerfect 2.1  
; CURRENT APPLICATION DATA: US/07977,434  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 590,490  
; FILING DATE: 28-SEP-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 590,466  
; FILING DATE: 28-SEP-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 590,213  
; FILING DATE: 28-SEP-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 523,394  
; FILING DATE: 15-MAY-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 143,441  
; FILING DATE: 12-JAN-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 063,509  
; FILING DATE: 17-JUN-1987  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 899,241  
; FILING DATE: 22-AUG-1986  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 746,121  
; FILING DATE: 15-AUG-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/US90/07641  
; FILING DATE: 21-DEC-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 585,471

; FILING DATE: 20-SEP-1990  
; PRIOR APPLICATION DATA: US 455,611  
; FILING DATE: 22-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 609,157  
; FILING DATE: 02-NOV-1990  
; PRIOR APPLICATION DATA: US 557,517  
; APPLICATION NUMBER: 24-JUL-1990  
; FILING DATE: 24-JUL-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Luann Cseri  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: Case No. 5466591 8753  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 814-2972  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 892 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-977-434-12  
  
Query Match 3.28; Score 116; DB 1; Length 892;  
Best Local Similarity 19.28; Pred. No. 0.095;  
Matches 118; Conservative 113; Mismatches 229; Indels 156; Gaps 30;  
  
Qy 85 YENCLOALADVEVKYIVERTMLDPQHVSDKEV-----RAASTEADKRLSRFDIEMS 137  
Db 31 HTNAVYGLTKMLIRLKE-----HISICKDACVFLDSKGGSKKRDILETKANRP 82  
Qy 138 MRGD-IFERIVHLOETCD-LGKIKPEARRYLEKSIKMGKRNGLHLPEQVQNEIKSMKRM 195  
Db 83 STPDLLLEQIPYVEELVDALG-----IKVLKIEGF-----EADDIIATLSKFF 125  
Qy 196 SELCIDFNKNLEDDTFLVFSKAEALGALPDD--FIDSLEK--TD---DDKYITILKYPHY 248  
Db 126 ES---DFEK-----VNIITGDKLLQVSDKVFVVRVERGITDLVYDRNKNVIEKGIY 176  
Qy 249 FPMVKKCCIPETRRMEMAFNTRCKEENTIILOQL-----LPLRTKVKAKLLGYSTHA 300  
Db 177 -----PE-----QFKDYLSDVGDQIDNIPGVKGIGKKTAVSLLKKYNSLE 216  
Qy 301 DFVLEMTAKSTSRVTAFLDLSQKLK---PLGEAERFILMLKKKECKDRGFEDGKIN 357  
Db 217 NVLKNINLL--TEKLRLLEDSDKEDLQKSIELVELIYDVPMDVEKDEIIRGYNPDKLLK 274  
Qy 358 AWDLYYYMTQTEELKYSIDQEFLEKY-----PPIEVVTBGLLNT 396  
Db 275 VLKKEFFSSIIKEL--NLQEKLEKEYILVDNEDKLLKLAEEKEYKTKTSIDTETSLDPF 332  
Qy 397 YQELGLSFEQMTD-----AHVWNKSVTLTYVKDKATGEVLGQFYLDLYPREGKYNHA 449  
Db 333 EAKLVGISISTMEGKAYIPVSHFGAKNISKSLI--DKELQILQEKDYNIVQNLKFDYE 391  
Qy 450 --ACFGLQPG-----CLLPD-----GSRMMAVALVNVFQPVAG 482  
Db 392 IFKSMGFSPNVPFHDTMTAAAYLLNPDKFRNLEELSLKYLKYMISFDEL-VNENVPFLG 450  
Qy 483 RP-SLLRHDEVITYPHEFGVHMQICAQTDFAFSGT-----NVETDFVEYPSQMLENW 535  
Db 451 NDFSVPPLERAVEYSCEDADVTRIFRLGRKIYENEMEKLFYEIMPLDVLSEMLNG 510  
Qy 536 VW-DVDSLRLRLSKHYKDGSPIDDLLEKL--VASRLVNTGLLT-LRQIVLSKVDOSLHTN 591  
Db 511 VFDEEYELKELSKYQEK---MDGIKEKVEIAGETFNLSSTQVAYILFEKINIAPYKK 567  
Qy 592 TSLDAASEYAKYCSFI 607  
Db 568 TATGKFSTNAEVLEEL 583

RESULT 9  
US-08-458-819-12  
Sequence 12, Application US/08458819  
Patent No. 5795762  
GENERAL INFORMATION:  
APPLICANT: Gelfand, David H.  
APPLICANT: Abramson, Richard D.  
TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF  
TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann-La Roche Inc.  
STREET: 340 Kingsland Street  
City: Nutley  
STATE: New Jersey  
ZIP: 07110-1199  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: 7  
SOFTWARE: WordPerfect 2.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/458,819  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/977,434  
FILING DATE: 23-FEB-1993  
APPLICATION NUMBER: US 590,490  
FILING DATE: 28-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 590,466  
FILING DATE: 28-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 590,213  
FILING DATE: 28-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 523,394  
FILING DATE: 15-MAY-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 143,441  
FILING DATE: 12-JAN-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 063,509  
FILING DATE: 17-JUN-1987  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 899,241  
FILING DATE: 22-AUG-1986  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 746,121  
FILING DATE: 15-AUG-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US90/07641  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 585,471  
FILING DATE: 20-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 455,611  
FILING DATE: 22-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 609,157  
FILING DATE: 02-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 557,517  
FILING DATE: 24-JUL-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Luann Cserr  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: Case No. 5795762 8753  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 814-2972

INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 892 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-458-819-12  
Query Match 3.28; Score 116; DB 1; Length 892;  
Best local similarity 19.24; Pred. No. 0.095;  
Matches 118; Conservative 113; Mismatches 229; Indels 156; Gaps 30;  
QY 85 YENCLOALADVEVKYIVERTMLDPQHVSSDKV-----RAASTADKRLSRFDIEMS 137  
DB 31 HTNAVYGLTKMLIKFLAE-----HISICKDACVFVLDKSGSKRKKDILEYKANRP 82  
QY 138 MRGD-IFERIVHLOETCD-LGKIKPEARLYLEKSIKMGKRNGLHLPEOVQNEIKSMKRM 195  
DB 83 STPDLLLEQIPYVEELVDALG-----IKVLKIEGF---EADDIIATLSKFK 125  
QY 196 SELCIDENKNLNEDDTFLVFSKALGALPDD--FIDSLEK--TD---DDKYKITLKYPHY 248  
DB 126 ES----DFEK-----VNITGDKDLQLVSKVFWVRVERGITDLVLYDRNKVIEKYIY 176  
QY 249 FPMVKKCCIPETRRRMEAFNTRCKEENTIILOOL-----LPLRTKVKAKLLGYSTHA 300  
DB 177 -----PE-----QFKDYLSLVGDQIDNIPGVKGIGKKTAVSLLKKYNSLE 216  
QY 301 DFVLEMTAKSTSRVTAFLDLSQKLK---PLGEAREFEFLNKKKCKEKGDFYDQKIN 357  
DB 217 NVLKNINLL--TEKLRLLEDSKEDLQKSIELVELIVDPMDVEKDEIIRGYNPKLLK 274  
QY 358 AWDLIYMTQTEELKYSIDQFLKEY-----FPIEVVTEGLLNT 396  
DB 275 VLKKEFESSIIEKEL--NLQEKLEKEYILVDNEDKLLKLAEEIEKYKTFSDITETSLDPF 332  
QY 397 YQELLGLSFEQMTD-----AHVMNKSVTLYTKDKATGEVLGQFYLDLPREGKYNHA 449  
DB 333 EAKLVGISISTMECKAYIIPVSHEFGAKNISLSLI-DKFLKQILQEKDYNIVGQNLKFDYE 391  
QY 450 --ACFGLQPG-----CLLPD-----GSRMMAVAALVYVNFSPQVAG 482  
DB 392 IFKSMGFSPNVPHPDITMAAYLLNPDEKRFNLEELSLKYLGYKMKISFDEL-VNENVPFLG 450  
QY 483 RP-SLLRHDEVRTYFHEFGVHWHQICATDFARESGT-----NVETDFVEVPQOMLENW 535  
DB 451 NDFSIVPLERAVEYSCEDADVTYRFRKLGRKIYENEMEKLFYEIEMPLDVLSEMEUNG 510  
QY 536 VW-DVDSLRLRSKHYKDGSPITADDLLEKL--VASRLVNTGLLT-LRQIVLSKVQDSLHTN 591  
DB 511 VYFDEEVLKLSKKYQEK---MDGIKEKVFETACETFNLSSTQVAYILFEKLIAPYKK 567  
QY 592 TSLDAASEYAKYCSEI 607  
DB 568 TATGKFSTNAEVLLEL 583  
RESULT 10  
PCT-US91-07035-12  
Sequence 12, Application PC/TUS9107035  
GENERAL INFORMATION:  
APPLICANT: Gelfand, David H.  
APPLICANT: Abramson, Richard D.  
TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF  
TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cetus Corporation  
STREET: 1400 Fifty-third Street  
City: Emeryville  
STATE: California  
ZIP: 94608

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Wordperfect 5.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/07035  
FILING DATE: 19910930  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 590,490  
FILING DATE: 28-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 590,466  
FILING DATE: 28-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 590,213  
FILING DATE: 28-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 523,394  
FILING DATE: 15-MAY-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 143,441  
FILING DATE: 12-JAN-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 063,509  
FILING DATE: 17-JUN-1987  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 899,241  
FILING DATE: 22-AUG-1986  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 746,121  
FILING DATE: 15-AUG-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US90/07641  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 585,471  
FILING DATE: 20-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 455,611  
FILING DATE: 22-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 609,157  
FILING DATE: 02-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 557,517  
FILING DATE: 24-JUL-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Sias Ph.D, Stacey R.  
REGISTRATION NUMBER: 32,630  
REFERENCE/DOCKET NUMBER: Case No. 2580  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-420-3300  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 892 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US91-07035-12

Query Match 3.2%; Score 116; DB 5; Length 892;  
Best Local Similarity 19.2%; Pred. No. 0.095;  
Matches 118; Conservative 113; Mismatches 229; Indels 156; Gaps 30;

QY 85 YENCLOALADVEVKIVERTMLDFQHVSSKEV-----RAASTEADKRLSRFDIEMS 137  
DB 31 HTNAVYGLTKMLIKFLKE-----HISIGKDACVFVLDKSGSGKKRDKDILETYKANRP 82  
QY 138 MRGD-IFERIVHLOETCD-LGKIKPEARRYLEKSIKMGKRNGLHLPQVQNEIKSMKKRM 195

Db 83 STPDLLLEQIPYVEELVDALG-----IKVLKIEGF-----EADDIIATLSKKF 125  
QY 196 SELCIDFNKNLNEDDTFLVFSKAGLALPD--FIDSLEK--TD---DDKYKITLKYPHY 248  
Db 126 ES---DFEK-----VNIITGDKDLQLSDKVFVWRVERGITDLVLVDRKNVIEKYGIY 176  
QY 249 FPMVKCCIPETRRRMMEMAFNTRCKEENTIILOQL-----LPLRTKVAKLLGYSTHA 300  
Db 177 -----PE-----QFQDLSLVGDQIDNIPGVKGICKKTAVSLLKKYNSLE 216  
QY 301 DFVLEMTAKTSRVTAFPLDDLSQKLK---PLGPAEREFILNLKKKECKDRGFEDYDGIN 357  
Db 217 NVLKNINLL--TEKLRRLLEDSDKEDLQKSIELVELLYDVPMDVEKDELIYRGYNPKLLK 274  
QY 358 AMDLYYMTQTEELKYSIDQEFLEKY-----PPIEVVTEGLLNT 396  
Db 275 VLKKEFSSIIKEL--NLQEKLEKEYILVDNEDKLLKLAEEIEKYKTSIDTETSLDPF 332  
QY 397 YOELLGLSFEQMTD-----AHWNKSVTLTYVKDKATGVFLGQFYLDLYPREGKYNHA 449  
Db 333 EAKLVGISISTMEGKAYIPVSHFGAKNISKSLI-DKFLKQILQEKDYNIVGONLKFDYE 391  
QY 450 --ACFGLQPG-----CLLPD-----GSRMMAVAALVVFNSQPVAG 482  
Db 392 IFKSMGFSPNVPFHEDTMTAAAYLLNPDEKRFNLEELSLKYLGYKMSFDEL-VNENVPFLG 450  
QY 483 RP-SLLRHDEVRTYFHERGHVHQAICATDFARSGT-----NVETDFVEVPSONLENN 535  
Db 451 NDFSVPPLERAVEYSCEDADVTYIFRKLKRIYENEMEKLFYEIEMPLIDVLSEMEJNG 510  
QY 536 VW-DVDSLRRLSKHYKDGSPITADLLEKL--VASRLVNTGLIT-LRQIVLSKVDQSLHTN 591  
Db 511 VYFDEEYLKLSKKYQEK--MDGIKEKVFETAGETFNLSSTQVAYILFEKLNIAPYKK 567  
QY 592 TSLDAASEYAKYCSEI 607  
Db 568 TATGKFSTNAEVLLEL 583

RESULT 11  
US-09-128-142-2  
; Sequence 2, Application US/09128142  
; Patent No. 6294559  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Roy G.  
; TITLE OF INVENTION: ANTIPROLIFERATIVE AGENTS ASSOCIATED WITH  
; PEROXISOME PROLIFERATOR ACTIVATED RECEPTORS GAMMAL AND G  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Jack L. Tribble  
; STREET: 126 E. Lincoln Ave., P.O. Box 2000  
; CITY: Rahway  
; STATE: New Jersey  
; COUNTRY: US  
; ZIP: 07065-0907  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: Power Macintosh 7500/100  
; SOFTWARE: Microsoft Word 6.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/128,142  
; FILING DATE: 03-Aug-1998  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/844,007  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Tribble, Jack L.  
; REGISTRATION NUMBER: 32,633  
; REFERENCE/DOCKET NUMBER: <Unknown>  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (908) 594-5321



APPLICANT: Davis, Maria  
APPLICANT: Sha, Dan  
TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASE  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
SUITE: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/766,014  
FILING DATE: Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below  
APPLICATION NUMBER: 60/008,688  
FILING DATE: December 15, 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 223/104  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 872 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-766-014-2

Query Match 3.1%; Score 115; DB 1; Length 872;  
Best Local Similarity 17.6%; Pred. No. 0.11;  
Matches 108; Conservative 115; Mismatches 195; Indels 196; Gaps 27;  
Qy 54 PEQKTRTELIVQKYDVGMLGIEEVTY-----NCIQALAD---VEVKYIV-ERT 104  
Db 80 PEELAEQVDYL---KEIIDGNFKITLEGEYADDDIIGTISKLAEEKGMEVLVVTGDRD 135  
Qy 105 MLDPEQVSSDKVRAASTADKRLSRDIEMSRGDIFERI-VHLQETCDLGKIKPEAR 163  
Db 136 AL--QVSDKVKIKTSKKGITQMEFEDEKA-----ILERYGITPOQFIDLGUMGDKS 186  
Qy 164 RYLEKSIKMKRNGHL-----PEQVQN--EIKSMKKRMSELC- 199  
Db 187 DNIQVNPNGEIKTAIKLLKDFGTIENLQNLSQLKGIKENIENKLAISKRLATIKR 246  
Qy 200 -----IDFN-----KNLNEDDTFLVFSKAEALPDFTDLSLEKTDKDKYKTLKYPHYFP 250  
Db 247 DIPTIEDPEEYKVKFNEEKLLELFNKLFEFSL-----IDNIKESS----- 288  
Qy 251 VMKKCCIPETRRRMEAFNRCKEENTIILOQLPLRTPKVAKLLGYSTHADFVLEMTAK 310  
Db 289 -----TEIVDNHVKVSKVVDIKELVTLQDNRNIAFPYLIYEIEIK-KIAF 334  
Qy 311 STSRVTAFLDDL-SQKPLGEAE-----REFINLKKK--ECKDRGFEDGKINA 358  
Db 335 SFGKDTVIDVFQEDLKEIFEKEDFEFTTHEIKDFLVLSYKGLCKSK--YIDTAVNA 392  
Qy 359 WDLYYMTQTTELKYSIDQEFLEKFFPIEVVVT-EGLLNTYOELLGLS--FQGMTDAHVN 415

Db 393 -----YLLNPSESNDLDR-VLKYLKVDVPSYEGIFGKGRDKKKIEIDENILADYICS 446  
Qy 416 KSVTLTYTKDKATGEVLGQFYLDLYPREGKYNHAACFGLOPGCLLPDGSRMVAALVWN 475  
Db 447 KCVYFLDKKEK-----LMNFIEMDMKKLLLE 473  
Qy 476 FSQPVAGRPSLLRHRDEVITYFHEFGVHMHQICAOQDFARFSGTNTVETDFVEVPSQMLEN- 534  
Db 474 IEMPLV---EVLKSMEV-----SGFTLDKEVLKELSKQIDDR 507  
Qy 535 -----WVVDVDSLRRLS-----KHFKDGSPIADDLLEKLVASRLV 569  
Db 508 IGEILDKIYKEAGYQFNWNSPKQLSEFLFEKLNLPVIAKTKTKYSTDSEVLEQLVPYNDI 567  
Qy 570 NTGLLTQLRQIVLSK 583  
Db 568 VSDIIEYRQLTKLK 581  
RESULT 14  
US-08-245-511-47  
Sequence 47, Application US/08245511  
Patent No. 5928900  
GENERAL INFORMATION:  
APPLICANT: Masure, H Robert  
APPLICANT: Pearce, Barbara J  
APPLICANT: Tuomanen, Elaine  
TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND  
TITLE OF INVENTION: ACCELLULAR VACCINES BASED THEREON  
NUMBER OF SEQUENCES: 58  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/245,511  
FILING DATE: 01-SEP-1994  
FILING DATE: 18-MAY-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/116,541  
FILING DATE: 01-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-069 CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 47:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 643 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-245-511-47

Query Match 3.1%; Score 113.5; DB 2; Length 643;  
Best Local Similarity 19.4%; Pred. No. 0.095;  
Matches 86; Conservative 57; Mismatches 152; Indels 149; Gaps 16;  
Qy 296 YSTHADFVLEMTAK-STSRVTAFLDDLSQLKPLGEAREFEILNKKKECKDRGFEDG 354

```
Db 23 YETDPNLNLYLTAKAANTANITSNVVD-----GLLENDRYGNFVPSMAEDWSVSKDG 74
QY 355 KINAW-----DLYYMTQTEE-----LKYSIDQEFLEKYEPIEVVTEGLLNTYQ 398
Db 75 LYYTIRKDAKWYTSEGEYAAVKAQDFVTGLKYAADKK--SDALYPVOESIKGL----- 128
QY 399 ELLGLSFEQMTDAHVNKSVTLTYTKDKATGEVLGQFYLDLYPREGKYNHAAACFGL----- 454
Db 129 -----DAYVKEIKDFSQVGIKALDEQTVQYTLN--KPEFWNSKTTMGVLAPV 175
QY 455 -----QPCCLLPDGSRRMVAALVYNFESQPVAGRPSLLRHDEVRTYFHE 498
Db 176 NEEFLNSKGGDDFAKATDPSSLLYNGPYLL--KSIYTKSSVEFAKNPNYWDKD----- 225
QY 499 FGHVMHQICAQTDFAFSGTNETDFEVEYPSQMLENWNWVDVSLRLSKHYKDGSPDAD 558
Db 226 -----NVHIDKVKLS-----FWDGQDTSKPAENFKDGSILTAAR 258
QY 559 L-----LEKLVASRLVNTGLLTLRQIVLSKVDQSLHTNTSLDAASEYAK----- 602
Db 259 LYPTSASFAELEKSMKDNIVYTQDSITYLVGTNIDROSQYKTSKTSDBQKASTKKALLN 318
QY 603 -----YCSILGVAATP-----GTNMPATFGLAGGYDGY-----YGY 636
Db 319 KDFRQAIAGFDRTAYASQLNGTGASKILRLNLFVPTFVQADGKNFGDMVKEKLVTYGD 378
QY 637 LMSEVFSMDMFYSCFKKEGIMNPE 660
Db 379 EWKDVNLAD-----SQDGLYNPE 396
```

RESULT 15

```
US-08-600-993A-47
; Sequence 47, Application US/08600993A
; Patent No. 5981229
; GENERAL INFORMATION:
; APPLICANT: Masure, H Robert
; APPLICANT: Pearce, Barbara J
; APPLICANT: Tuomanen, Elaine
; TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND
; TITLE OF INVENTION: ACCELLULAR VACCINES BASED THEREON
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/600,993A
; FILING DATE: 1-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/116,541
; FILING DATE: 01-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-069 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
```

```
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 643 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-600-993A-47

Query Match 3.1%; Score 113.5; DB 2; Length 643;
Best Local Similarity 19.4%; Pred. No. 0.095;
Matches 86; Conservative 57; Mismatches 152; Indels 149; Gaps 16;

QY 296 YTHADFLVEMTAK--STSRVTAFLDLDLSQKLPLCEAEREFLNLKKKECKDRGEYDG 354
Db 23 YETDPNLNLYLTAKAANTANITSNVVD-----GLLENDRYGNFVPSMAEDWSVSKDG 74
QY 355 KINAW-----DLYYMTQTEE-----LKYSIDQEFLEKYEPIEVVTEGLLNTYQ 398
Db 75 LYYTIRKDAKWYTSEGEYAAVKAQDFVTGLKYAADKK--SDALYPVOESIKGL----- 128
QY 399 ELLGLSFEQMTDAHVNKSVTLTYTKDKATGEVLGQFYLDLYPREGKYNHAAACFGL----- 454
Db 129 -----DAYVKEIKDFSQVGIKALDEQTVQYTLN--KPEFWNSKTTMGVLAPV 175
QY 455 -----QPCCLLPDGSRRMVAALVYNFESQPVAGRPSLLRHDEVRTYFHE 498
Db 176 NEEFLNSKGGDDFAKATDPSSLLYNGPYLL--KSIYTKSSVEFAKNPNYWDKD----- 225
QY 499 FGHVMHQICAQTDFAFSGTNETDFEVEYPSQMLENWNWVDVSLRLSKHYKDGSPDAD 558
Db 226 -----NVHIDKVKLS-----FWDGQDTSKPAENFKDGSILTAAR 258
QY 559 L-----LEKLVASRLVNTGLLTLRQIVLSKVDQSLHTNTSLDAASEYAK----- 602
Db 259 LYPTSASFAELEKSMKDNIVYTQDSITYLVGTNIDROSQYKTSKTSDBQKASTKKALLN 318
QY 603 -----YCSILGVAATP-----GTNMPATFGLAGGYDGY-----YGY 636
Db 319 KDFRQAIAGFDRTAYASQLNGTGASKILRLNLFVPTFVQADGKNFGDMVKEKLVTYGD 378
QY 637 LMSEVFSMDMFYSCFKKEGIMNPE 660
Db 379 EWKDVNLAD-----SQDGLYNPE 396
```

Search completed: January 14, 2002, 18:56:33  
Job time: 792 sec

---